Polypeptides implicated in the expression of resistance to glycopeptides, in particular in Gram-positive bacteria. Nucleotide sequence coding for these polypeptides and use for diagnosis

The invention relates to the polypeptides associated with the expression of resistance to antibiotics of the glycopeptide family, in particular in Gram-positive bacteria, in particular in the family of the Gram-positive cocci. The invention also relates to a nucleotide sequence coding for these polypeptides. It also relates to the use of these polypeptides and their nucleotide sequence as agents for the in vitro detection of resistance to glycopeptides. Among the Gram-positive cocci, the invention relates most particularly to the enterococci, the streptococci and the staphylococci which are of particular importance for the implementation of the invention.

The glycopeptides, which include vancomycin and teicoplanin are antibiotics which inhibit the synthesis of the bacterial cell wall. These antibiotics are very much used for the treatment of severe infections due to Gram-positive cocci (enterococci, streptococci and staphylococci), in particular in the light of allergy and resistance to the penicillins. In spite of long clinical usage of vancomycin, this antibiotic has remained active towards almost all of the strains up to 1986, the date at which the first resistant strains were isolated. Since then, resistance to the glycopeptides has been detected by many microbiologists in Europe and in the United States, in particular in strains isolated from immunodepressive patients, making necessary a systematic evaluation of the sensitivity of the microbes in hospital environments.

The activity of the glycopeptides depends on the formation of a complex between the antibiotic and the precursors of the peptidoglycan, more than on the direct interaction with enzymes of cell wall metabolism. In particular, it has been observed that the glycopeptides bind to the terminal D-alanyl-D-alanine residues (D-ala-D-ala) of the precursors of the peptidoglycan.

The recent emergence of resistance to the glycopeptides, in particular in the enterococci, has led to certain results being

obtained with regard to knowledge of the factors conferring this resistance.

For example it has been observed in a particular strain of enterococci, Enterococcus faecium BM4147, that the determinant of resistance to the glycopeptides is localized on a plasmid of 34 kb, the plasmid pIP816. This determinant has been cloned in E.coli (Brisson Noel et al., 1990, Antimicrob Agents Chemother 34, 924-927).

According to the results hitherto obtained, the resistance to the glycopeptides is associated with the production of a protein of molecular weight of about 40 kDa, the synthesis of this protein being induced by sub-inhibitory concentrations of certain glycopeptides such as vancomycin.

By carrying out a more detailed study of the resistance of certain strains of Gram-positive cocci towards glycopeptides, in particular vancomycin or teicoplanin, the inventors have observed that this resistance might be linked to the expression of several proteins or polypeptides encoded in sequences usually borne by plasmids in the resistant strains. The recent results obtained by the inventors also make it possible to distinguish the genes coding for two phenotypes of resistance, on the one hand strains highly resistant to the glycopeptides, and, on the other, strains with a low level of resistance.

By strain with a high level of resistance is meant a strain of bacteria, in particular a strain of Gram-positive cocci, for which the minimal inhibitory concentrations (MIC) of vancomycin and teichoplanin are higher than 32 and 8 µg/ml, respectively. The MIC of vancomycin towards strains with low-level resistance are included between 16 and 32 µg/ml. These strains are apparently sensitive to teicoplanin.

The inventors have isolated and purified, among the components necessary for the expression of the resistance to the glycopeptides, a particular protein designated VANA or VanA which exhibits a certain homology with D-alanine-D-alanine ligases. VanA is nonetheless functionally distinct from the ligases.

In principle, a gene sequence will be designated by "van..."

and an amino acid sequence by "Van..."

The invention relates to polypeptides or proteins implicated in the expression of resistance to antibiotics of the glycopeptide family and, in particular, to vancomycin and/or teicoplanin as well as to the nucleotide sequences coding for such complexes.

The invention also relates to nucleotide probes which can be used for the detection of resistance to the glycopeptides, in particular by means of the polymerase chain reaction (PCR), or by tests involving antibodies.

The invention relates to a composition of polypeptides, characterized in that it contains at least one protein or part of a protein selected from the amino acid sequences identified in the list of the sequences as ~~SED~~ SEQ ID NO 2 (VanH), SEQ ID NO 4 (VanA), SEQ ID NO 6 (VanX) or SEQ ID NO 8 (VanC), or any protein or part of a protein recognized by the antibodies directed against VanH, VanA, VanX or VanC, or any protein or part of a protein encoded in a sequence hybridizing with one of the nucleotide sequences identified in the list of the sequences as SEQ ID NO 1, SEQ ID NO 3, SEQ ID NO 5 or SEQ ID NO 7 or with one of the following sequences V1 (SEQ ID NO:9) or V2 (SEQ ID NO:10) under stringent or only slightly stringent conditions:

V1 : GGX GAA GAT GGX TCX TTX CAA GGX
G C AG C G
A

V2 : AAT ACX ATX CCX GGX TTT AC
C T C
C

A first particular composition according to the invention implicated in the expression of the resistance to the glycopeptides is characterized in that it comprises at least 3 proteins or any part of one or more of these proteins necessary to confer to Gram-positive bacteria the resistance to antibiotics of the glycopeptide family, in particular to vancomycin and/or teicoplanin or to promote this resistance, in particular in strains of the family of the Gram-positive cocci, these proteins or parts of proteins being

a) recognized by antibodies directed against one of the sequences identified in the list of the sequences as SEQ ID NO 1, SEQ ID NO 2, SEQ ID NO 3,

b) or encoded in genes containing a sequence identified as SEQ ID NO 4, SEQ ID NO 5 or SEQ ID NO 6 or hybridizing with one of these sequences or its complementary sequence or with the sequences V1 (SEQ ID NO:9) or V2, (SEQ ID NO:10) under stringent or only slightly stringent conditions.

These sequences are also designated, respectively, by ORF3, ORF1 containing the gene VanH, vanA (or ORF2); they characterize the proteins responsible for resistance as obtained from the strain Enterococcus faecium BM4147 described by Leclerq et al (N. Engl. J. Med. 319:157-161).

Another protein, VanC, (SEQ ID NO:8) related to the D-Ala-D-Ala ligases but of different specificity has been characterized in Enterococcus gallinarum BM4173; the vanC gene (SEQ ID NO:7) possesses domains having sufficient homology with the vanA gene for probes corresponding to defined regions of vanA to make possible its detection.

E.gallinarum is a constitutive isolate resistant to low levels of vancomycin (Dutka-Malen et al., Antimicrob. Agents Chemother 34 (1990b) 1875-1879).

By the expression "polypeptides" is meant any sequence of amino acids constituting proteins or being of a size less than that of a protein.

The stringent conditions mentioned above are defined according to the usual conditions pertaining to the hybridization of nucleotide sequences. As an example, in the case of the sequences which hybridize with the sequence of the vanA gene (SEQ ID NO 1) it will be possible to apply the following conditions:

- for hybridization under conditions of high stringency:
  * a reaction temperature of 65°C overnight in a solution containing 0.1% SDS, 0.7% skimmed milk powder, 6xSSC (1xSSC = 0.15 M NaCl and 0.015 M sodium citrate at pH = 7.0)
  * washes at 65°C in 2xSSC - 0.1% SDS;
- for hybridization under slightly stringent conditions, the hybridization temperature is 60°C overnight and the temperature

of the washings is 45°C.

The expression of resistance to glycopeptides may be expressed by the persistence of an infection due to microbes usually sensitive to the glycopeptides.

A polypeptide or a protein is necessary for the expression of resistance to the glycopeptides, if its absence makes the strain which contains this polypeptide or this protein more sensitive to the glycopeptides and if this polypeptide or protein is not present in sensitive strains.

Different levels of resistance to the glycopeptides exist in the strains of Gram-positive cocci, in particular.

According to a preferred embodiment of the invention, the polypeptides included in the composition defined above correspond to the combination of the proteins identified in the list of the sequences as SEQ ID NO 2 (VanH), SEQ ID NO 4 (VanA), SEQ ID NO 6 (VanX).

The inventors have thus observed that the expression of resistance to the glycopeptides in Gram-positive bacteria requires the expression of at least three proteins or of polypeptides derived from these proteins.

According to a first particular embodiment of the invention, the polypeptides of the composition are also characterized in that the amino acid sequences necessary for the expression of resistance to antibiotics of the glycopeptide family are under the control of regulatory elements, in particular of the proteins corresponding to the sequences designated by SEQ ID NO 12 and SEQ ID NO 14 in the list of the sequences, and which correspond to a regulatory sequence R and to a sensor sequence S, respectively.

VanS and VanR constitute a two-component regulatory system, VanR being an activator of transcription and VanS stimulating the transcription dependent on VanR. VanS is capable of modulating the level of phosphorylation of VanR in response to the vancomycin present in the external medium and is thus involved in the control of the transcription of the genes for resistance to vancomycin.

These regulatory sequences are in particular capable of increasing the level of resistance, to the extent to which they promote

the expression of the proteins responsible for resistance comprised in the polypeptides of the invention.

According to another advantageous embodiment of the invention, the polypeptides of the above composition are encoded in the sequence SEQ ID NO 15 identified in the list of the sequences, which represents the sequence coding for the 5 proteins previously described.

Another sequence according to the invention is designated by SEQ ID NO 16 which contains the sequence SEQ ID NO 15 as well as a sequence upstream from SEQ ID NO 15 coding for a transposase (encoded in the (-) strand of the sequence, and a sequence downstream from SEQ ID NO 6 corresponding to the genes vanY and vanZ and at each end reverse repeated sequences of 38 bp. SEQ ID NO 16 constitutes a transposon, the genes of which are implicated at different levels in the establishment of resistance to the glycopeptides.

The invention also relates to the purified proteins belonging to the composition and to the polypeptides described previously. In particular, the invention relates to the purified protein VanA, characterized in that it corresponds to the amino acid sequence SEQ ID NO 4 in the list of the sequences or a protein VanC, encoded in a gene capable of hybridizing with the vanA gene.

The protein VanA contains 343 amino acids and has a calculated molecular mass of 37400 Da. The protein VanC contains 343 amino acids and has a calculated molecular mass of 37504 Da.

Other interesting proteins in the framework of the invention correspond to the sequences identified as SEQ ID NO 2 (VanH), SEQ ID NO 6 (VanX), SEQ ID NO 12 (VanR), SEQ ID NO 14 (VanS) in the list of the sequences.

The sequence identified by the abbreviation SEQ ID NO 1 contains the protein VanH encoded in the gene vanH, this protein contains 322 amino acids and begins with a methionine. This protein is an enzyme implicated in the synthesis of the peptidoglycan and has a molecular mass of 35,754 kDA. VanH exhibits some similarities to dehydrogenases which catalyze the NAD^+-dependent oxidation of 2-hydroxy-carboxylic acids to form the corresponding 2-keto-carboxylic acids. In fact, the VanH protein might use $NADP^+$ rather than NAD^+. The VanH

protein also contains several residues of reactive sites which probably participate directly in the binding of the substrate and in catalysis. VanH might be implicated in the synthesis of a substrate of the ligase VanA. This substrate of VanA might be a D-α-hydroxy-carboxylic acid, which might be condensed by VanA with D-alanine in the place of a D-amino acid, which might affect the binding of the precursor of the peptidoglycan with vancomycin, as a result of the loss of a hydrogen bond because one of the hydrogen bonds formed between vancomycin and N-acetyl-D-Ala-D-Ala occurs with the NH group of the terminal D-alanine residue. Let it be recalled that "Ala" is the abbreviation for "alanine".

The inventors have been able to detect some interactions between the proteins VanA and VanH and have in particular been able to describe the following : the nature of the VanA protein (D-alanine: D-alanine ligase with reduced specificity for its substrate) which has made possible resistance to glycopeptides, implies the biosynthesis by VanA of a novel compound different from D-Ala-D-Ala, a peptide which may be incorporated into the peptidoglycans but which is not recognized by vancomycin. In particular, the observation of similarities between the product of the vanH gene and the D-specific α-keto-acid reductases has made it possible to determine that this compound cannot be a D-amino acid but is a D-hydroxy acid, which when it is bound to D-alanine by VanH, can generate the novel depsipeptide precursor of the peptidoglycan.

The invention also relates to any combination of these different proteins in a resistance complex, as well as to hybrid proteins comprising one or several of the above proteins, or part of these proteins, in combination with a defined amino acid sequence.

Also included in the framework of the invention are nucleotide sequences coding for one of the amino acid sequences described above.

A particular sequence is the nucleotide sequence of about 7.3 kb, corresponding to the HindIII-EcoRI restriction fragment, such as that obtained starting from the plasmid pIP816 described in the publication of Leclerq et al - 1988, cited above.

This sequence of 7.3 kb comprises the nucleotide sequence

coding for the 3 resistance proteins and the 2 regulatory proteins referred to above. This coding sequence is included in an internal BglII-XbaI fragment. It also comprises a part of the sequences coding for the transposase and the resolvase.

The invention also relates to any nucleotide fragment comprising the above-mentioned restriction fragment as well as any part of the HindIII-EcoRI fragment, in particular the EcoRI-XbaI fragment of about 3.4 kb coding for the 3 resistance proteins or the EcoRV-SacII fragment of about 1.7 kb coding for VanA or also HindIII-EcoRI fragment of about 3.3 kb coding for the 2 regulatory proteins VanR and VanS.

Another definition of a nucleotide sequence of the invention corresponds to a nucleotide fragment containing the following restriction sites in the following order, such as obtained starting from pIP816 mentioned above:

HindIII, BglII, BglII, EcoRI, BamHI, XbaI, EcoRI.

Another nucleotide sequence according to the invention is characterized in that it corresponds to a sequence selected from the sequences identified as SEQ ID NO ~~7~~ 15, SEQ ID NO ~~6~~ 17, ~~SEQ ID NO 11~~ or SEQ ID NO ~~22~~ 16, or in that it includes this sequence or any part of this sequence, or also any sequence or part of the sequence of the complementary DNA or any sequence of RNA corresponding to one of these DNAs, capable,

- either of constituting a hybridization probe for the detection of resistance to antibiotics of the glycopeptide family, in particular to vancomycin and/or teicoplanin in particular in strains of the family of the Gram-positive cocci,
- or of coding for a sequence necessary or associated with the expression of resistance to antibiotics of the glycopeptide family, in particular to vancomycin and/or teicoplanin., in particular in strains of the family of the Gram-positive cocci.

The sequence SEQ ID NO ~~7~~ 17 codes for the 3 resistance proteins VanH, VanA and VanX.

The sequence SEQ ID NO ~~22~~ 16 ~~and the sequence SEQ ID NO 11 include~~ includes a transposon shown in Figure 7a; this transposon contains the

genes necessary for the expression of resistance to the glycopeptides as well as the genes associated with this resistance implicated, for example, in the regulation of the expression of the genes necessary to produce the resistance phenotype or implicated in the amount of resistance polypeptide produced.

A specific sequence corresponding to the above definition is one of the following sequences:

```
     (SEQ ID NO:9)
V1 : GGX GAA GAT GGX TCX TTX CAA GGX
         G   C       AG  C       G

or                       A
     (SEQ ID NO:10)
V2 : AAT ACX ATX CCX GGX TTT AC
       C       T             T
               C
```

V1 and V2 make possible the constitution of probes, if necessary, in combination with other nucleotides, depending on the degree of specificity desired in order to detect vanA and vanC and may also be used as primers in polymerase chain reactions.

Other preferred nucleotide sequences are the sequences SEQ ID NO 8, SEQ ID NO 9, SEQ ID NO 10, SEQ ID NO 21, SEQ ID NO 12 (transposase), SEQ ID NO 13 (resolvase), SEQ ID NO 14 (vanY), SEQ ID NO 15 (vanZ), SEQ ID NO 23 (vanR), SEQ ID NO 24 (vanS) or a variant of one of these sequences provided that it codes for a protein having immunological and/or functional properties similar to those of the proteins encoded in the sequences SEQ ID NO 8 (vanA), SEQ ID NO 9 (vanH), SEQ ID NO 10 (vanX), or SEQ ID NO 21 (vanC), SEQ ID NO 12 (transposase), SEQ ID NO 13 (resolvase), SEQ ID NO 14 (vanY), SEQ ID NO 15 (vanZ), SEQ ID NO 23 (vanR), SEQ ID NO 24 (vanS) or in that it makes possible the detection of strains resistant to antibiotics of the glycopeptide family.

Variants include all of the fragments of the sequences having the following properties.

These sequences code for the resistance proteins VanH, VanA and VanX.

The nucleotide sequence designated by SEQ ID NO 8 corresponds to a DNA fragment of 1029 bp situated between the ATG codon at position

377 and the TGA codon at position 1406 on the plasmid pAT214 (Fig. 6).

The invention also relates to a nucleotide sequence coding for the sequence SEQ ID NO 15 corresponding to the sequence coding for the 5 proteins (2 regulatory proteins and 3 resistance proteins), and also comprising the flanking sequences associated with these coding sequences, or comprising this sequence.

Also included in the framework of the invention is a sequence modified with respect to SEQ ID NO 15, characterized in that it lacks the flanking sequences. These flanking sequences are the sequences shown in the following pages and defined as follows:

- sequence upstream from the sequence coding for R: between the bases 1 and 1476 of the sequence shown in Figure 5,
- sequence between the sequence coding for the sensor protein S and ORF1: between the bases 3347 and 3500 of the sequence shown in Figure 5,
- sequence downstream from the sequence coding for ORF3: between the bases 6168 and 7227 of the sequence shown in Figure 5.

The sequence designated by SEQ ID NO 15 is also characterized by the fragment bearing the restriction sites in the following order:

BglIII - EcoRI - BamHI - EcoRI

The location of the regulatory proteins and the resistance proteins is shown in Figure 3.

The inventors have identified upstream and downstream from the genes vanR, vanS, vanH, vanA and vanX, which are necessary for or associated with the expression of resistance to glycopeptides at a given level, genes coding for a transposase and a resolvase (upstream from the group previously mentioned) and genes vanY and vanZ, downstream from this group. The genes for the transposase and resolvase might be implicated in transposition functions and the vanY gene coding for a D,D-carboxy peptidase might be implicated in the metabolism of the peptidoglycan, and might contribute to resistance to the glycopeptides in E. faecium BM4147 even though vanR, vanS, vanH, vanA and vanX borne by a plasmid in a high number of copies, alone confer a high level of resistance.

Let it be noted that the sequence coding for the transposase- (SEQ ID NO: 18) is located on the (-) strand of the sequence ID NO 16 which codes for vanR, vanS, vanH, vanA, vanX, vanY, vanZ and the resolvase.

The invention relates not only to the DNA sequences identified in the list of the sequences but also to the complementary DNA sequences and the corresponding RNA sequences. The invention concerns in addition sequences which are equivalent to the former, either in terms of expression of proteins, polypeptides or their fragments described above, or in terms of the capacity to detect, for example by chain polymerization procedures, strains of Gram-positive bacteria exhibiting resistance to antibiotics of the glycopeptide family such as vancomycin or teicoplanin.

Recombinant sequences characterized in that they comprise one of the above nucleotide sequences also form part of the invention.

The invention also relates to a recombinant vector characterized in that it includes one of the above nucleotide sequences at a site inessential for its replication, under the control of regulatory elements likely to be implemented in the expression of the resistance to antibiotics of the glycopeptide family, in particular to vancomycin or teicoplanin. in a defined host.

Particularly advantageous recombinant vectors for the implementation of the invention are the following vectors: pAT214 containing the EcoRV-SacII fragment of 1761 bp containing a nucleotide sequence coding for the VanA protein; in these vectors the sequences of the invention are advantageously placed under the control of promoters such as the lac promoter.

The invention also relates to a recombinant cell host containing a nucleotide sequence such as that previously described or a vector such as that described above under conditions which make possible the expression of resistance to antibiotics of the glycopeptide family, in particular resistance to vancomycin and/or this host being for example selected from the bacteria, in particular the Gram-positive cocci.

In certain applications it is also possible to use yeasts, fungi, insect or mammalian cells.

The invention also relates to a nucleotide probe characterized in that it is capable of hybridizing with a sequence previously described, this probe being labelled if necessary. These probes may or may not be specific for the proteins of resistance to glycopeptides.

Labels which can be used for the requirements of the invention are the known radioactive labels as well as other labels such as enzymatic labels or chemoluminescent labels.

Probes thus labelled may be used in hybridization tests in order to detect resistance to glycopeptides in Gram-positive bacteria. In this case, conditions of low stringency will be used.

Nucleotide probes according to the invention may be characterized in that they are specific in Gram-positive bacteria for the sequences coding for a resistance protein to the glycopeptides, in particular to vancomycin and/or teicoplanin these probes being in addition universal among these sequences.

By these specific probes is meant any oligonucleotide hybridizing with a nucleotide sequence coding for one of the proteins according to the invention, such as described in the preceding pages, and not exhibiting a cross hybridization reaction or amplification reaction (PCR) with sequences present in all of the sensitive strains.

The universal character of the oligonucleotide which can be used in PCR is defined by their capacity to promote specifically the amplification of a nucleotide sequence implicated in resistance in any one strain of Gram-positive bacteria, resistant to the antibiotics of the glycopeptide family.

The size of the nucleotide probes according to the invention may vary depending on the use desired. For the oligonucleotides which are used in PCR, recourse will be had to fragments of a length which is usual in this procedure. In order to construct probes, it is possible to take any part of the sequences of the invention, for example probe fragments of 200 nucleotides.

According to a particular embodiment of the invention, a nucleotide probe is selected for its specificity towards a nucleotide sequence coding for a protein necessary for the expression in Gram-positive bacteria of a high level of resistance to antibiotics of the

glycopeptide family, in particular to vancomycin and teicoplanin.

As examples, useful probes may be selected from the intragenic part of the vanA gene.

Other useful probes for carrying out the invention are characterized by their universal character, according to the preceding definition, but are not specific for the resistance genes. They may also be used as primers in PCR, and are for example:

(SEQ ID NO:9)
V1 : GGX GAA GAT GGX TCX TTX CAA GGX
G C AG C G
A

(SEQ ID NO:10)
V2 : AAT ACX ATX CCX GGX TTT AC
C T C
C

V1 and V2 hybridize with vanA and vanC and are capable of leading to the detection of proteins associated with resistance to glycopeptides in other micro-organisms.

Other particular probes of the invention have the specific character of a nucleotide sequence coding for a protein necessary for the expression in Gram-positive bacteria of a low level of resistance to antibiotics of the glycopeptide family, in particular to vancomycin in Gram-positive bacteria.

It should also be mentioned that oligonucleotide probes which might be derived from the sequence of the vanA gene coding for the VanA protein may be used indiscriminantly to detect high-level or low-level resistance.

In a particularly preferred manner, a probe of the invention is characterized in that it hybridizes with a chromosomal or non-chromosomal nucleotide sequence of a Gram-positive strain resistant to glycopeptides, in particular to vancomycin and/or teicoplanin , in particular in that it hybridizes with a chromosomal or non-chromosomal nucleotide sequence of a strain of Gram-positive cocci, for example an enterococcal strain and preferably E. faecium 4147 or E. gallinarum.

In order to distinguish strains with a high level of resistance from strains with a low level of resistance it is possible to carry out a hybridization test using conditions of high stringency.

The oligonucleotides of the invention may be obtained from the sequences of the invention by cutting with restriction enzymes, or by chemical synthesis according to the standard methods.

Furthermore, the invention relates to polyclonal or monoclonal antibodies, characterized in that they recognize the polypeptide(s) described above or an amino acid sequence described above.

These antibodies may be obtained according to standard methods for antibody production. In particular, in the case of the preparation of monoclonal antibodies, recourse will be had to the method of Köhler and Milstein according to which monoclonal antibodies are prepared by cell fusion between myeloma cells and mouse spleen cells previously immunized with a polypeptide or a composition according to the invention, in conformity with the standard procedure.

The antibodies of the invention can advantageously be used for the detection of the presence of proteins characteristic of resistance to the glycopeptides, in particular to vancomycin and teicoplanin.

Particularly useful antibodies are polyclonal or monoclonal antibodies directed against the protein VanA or VanC. Such antibodies advantageously make it possible to detect strains of bacteria, in particular Gram-positive cocci, exhibiting high-level resistance to the antibiotics of the glycopeptide family. If necessary, a step entailing lysis of the cells of the sample undergoing detection is performed prior to the placing in contact of the sample with the antibodies.

In order to carry out this detection, recourse will advantageously be had to antibodies labelled for example with a radioactive substance or other type of label.

Hence, tests for the detection in Gram-positive bacteria of resistance to the glycopeptides, in particular tests making use of the ELISA procedures, are included in the framework of the invention.

A kit for the <u>in vitro</u> diagnosis of the presence of Gram-positive strains, resistant to the glycopeptides, in particular to vancomycin and/or teicoplanin, these strains belonging in particular to the Gram-positive cocci for example enterococci, for example <u>E.</u>

faecium or E. gallinarum is characterized in that it comprises:

- antibodies corresponding to the above definition, labelled if necessary,
- a reagent for the detection of an immunological reaction of the antigen-antibody type,
- if necessary, reagents to effect the lysis of the cells of the sample to be tested.

Furthermore, the agents developed by the inventors offer the very useful advantage of being suitable for the development of a rapid and reliable test or kit for the detection of Gram-positive strains resistant to the glycopeptides by means of the polymerase chain reaction (PCR). Such a test makes it possible to improve the sensitivity of the existing tests which remain rather unreliable and, in certain cases, may make possible the detection of all of the representatives of the family of the genes coding for resistance proteins to the glycopeptides in Gram-positive bacteria.

The carrying out of a test by means of the method of amplification of the genes of these proteins is done by the PCR procedure or by the RPCR procedure (RPCR : abbreviation for reverse polymerase chain reaction).

The RPCR technique makes possible the amplification of the NH_2 and COOH terminal regions of the genes it is desired to detect.

Some specific primers make it possible to amplify the genes of the strains with low-level resistance. These primers are selected, for example, from the sequence coding for the resistance protein VanA.

As examples, the following sequences can be used as primers for the preparation of probes for the detection of an amplification by means of the PCR or RPCR method.

V1 (SEQ ID NO:9) : GGX GAA GAT GGX TCX TTX CAA GGX
G C AG C G
A

V2 (SEQ ID NO:10) : AAT ACX ATX CCX GGX TTT AC
C T C
C

X represents one of the bases A,T,C or G or also corresponds in all cases to inosine.

Naturally, the invention relates to the complementary probes of the oligonucleotides previously described as well as possibly to the RNA probes which correspond to them.

A kit for the in vitro diagnosis of the presence of strains of Gram-positive bacteria resistant to the glycopeptides, in particular resistant to vancomycin and/or teicoplanin these strains belonging in particular to the Gram-positive cocci, in particular that they are strains of enterococci, for example E. faecium or E. gallinarum, is characterized in that it contains:

- a nucleotide probe complying with the above specifications and if necessary,
- oligonucleoside triphosphates in an amount sufficient to make possible the amplification of the desired sequence,
- a hybridization buffer,
- a DNA polymerization agent.

The invention also relates to a procedure for the in vitro detection of the presence of Gram-positive strains resistant to the glycopeptides, in particular to vancomycin and/or teicoplanin. these strains belonging in particular to the family of the Gram-positive cocci, in particular in that they are strains of enterococci, for example E. faecium or E. gallinarum, characterized in that it comprises:

a) the placing of a biological sample likely to contain the resistant strains in contact with a primer constituted by a nucleotide sequence described above, or any part of a sequence previously described, capable of hybridizing with a desired nucleotide sequence necessary for the expression of resistance to the glycopeptides, this sequence being used as matrix in the presence of the 4 different nucleoside triphosphates and a polymerization agent under conditions of hybridization such that for each nucleotide sequence which has hybridized with a primer, an elongation product of each primer complementary to the matrix is synthesized,

b) the separation of the matrix from the elongation product obtained, this latter then also being capable of behaving as a matrix,

c) the repetition of step a) so as to produce a detectable amount of the desired nucleotide sequences,

d) the detection of the product of amplification of the nucleotide sequences.

The detection of the elongation products of the desired sequence may be carried out by a probe identical with the primers used to carry out the PCR or RPCR procedure, or also by a probe different from these primers, this probe being labelled if necessary.

Details relating to the implementation of the PCR procedures may be obtained from the patent applications EP 0229701 and EP 0200362.

Other advantages and characteristics of the invention will become apparent in the examples which follow and from the figures.

FIGURES

- Figure 1 : electrophoresis on SDS-polyacrylamide gel (SDS-PAGE) of the proteins of the membrane fractions line 1 and line 4, molecular weight standards; line 2, E. faecium BM4147 placed in culture in the absence of vancomycin; line 3, BM4147 placed in culture in the presence of 10 µg/ml of vancomycin. The head of the arrow indicates the position of the VanA protein.

- Figure 2:

Figure2A

A : Restriction maps of the inserts of the plasmids pAT213 and pAT214. The vector and the DNA insert are distinguished by light and dark segments, respectively. The open arrow represents the vanA gene.

Figure2B

B : Strategy for the nucleotide sequencing of the insert of 1761 bp in the plasmid pAT214. The arrows indicate the direction and extent of the sequencing reactions by the dideoxy method. The synthetic oligonucleotide primer (5' ATGCTCCTGTCTCCTTTC 3' OH) (SEQ ID NO:27) is complementary to the sequence between the positions 361 and 378. Only the pertinent restriction sites are given.

- Figure 3 : position of the sequences R, S, ORF1, ORF2, ORF3.

- Figure 4 : representation of SEQ ID NO 15/6.

- Figure 5 : representation of SEQ ID NO 15/6 and the corresponding protein. (SEQ ID NOS:27, 28 AND 29)

- Figure 6 : sequence of the vanA gene and the corresponding protein.

- Figure 7 :

(a) (Figure 7A) : Localization of the genes vanR, vanS, vanH, vanA, vanX, vanY, vanZ of the gene for the transposase and of the gene for the resolvase as well as the repeated reverse terminal sequences of 38 bp at the end of the transposon.

(b) (Figure 7b) : Mapping of the plasmids. (A) Polylinker pAT29 and derivatives constructed in this study. The arrow labelled P2 indicates the position and orientation of the P2 promoter of aphA-3 (Caillaud et al., 1987, Mol. Gen. Genet. 207:509-513). (B) Insert pAT80. The white rectangles indicate the DNA of pAT29 but they are not shown to scale. The rectangles terminating in an arrow indicate the coding sequences. The arrows shown in vertical and horizontal full lines indicate the position and orientation, respectively, of the apha-1 gene in the derivatives of pAT80. Restriction sites: Ac, AccI; B, BamHI; Bg, BglII; Bs, BssHII; E, EcoRI; H, HindIII; Hc, HincII; K, KpnI; P, PstI; S, SmaI; SI, SacI, SII, SacII; Sa, SalI; Sp, SphI; Xb, XbaI. (C) Inserts in pAT86, pAT87, pAT88 and pAT89. The inserts are shown by full lines and the corresponding vectors are indicated in parentheses.

- Figure 8: nucleotide sequence of the transposon shown in Figure 7 (SEQ ID NOS:16 AND 30) and amino acid sequence of the corresponding proteins. The nucleotide sequence is shown for the (+) strand (SEQ ID NO:16) and for the (-) strand (SEQ ID NO:30) (corresponding to the complementary sequence of the (+) strand for the positions 1 to 3189) on which the coding sequence of the transposase is located.

- Figure 9 : Nucleotide sequence of the SacI-PstI fragment of 1347 bp of the plasmid pAT216 containing the vanC gene. (SEQ ID NO:31) The numbering starts

at the first base G of the SacI restriction site. The potential RBS sequence upstream from the initiation codon ATG of translation at position 215 is underlined. The STOP codon (TGA) is indicated by *. The region coding for the vanC and the deduced amino acid sequence are indicated in bold characters. Sequential overlapping clones were generated by restriction fragments of subcloning of pAT216 in the bacteriophage M13mp10 (Amersham, England). The universal primer (New England Biolabs Beverly MA) was used to sequence the insert in the recombinant phages. The sequencing was performed by the enzymatic dideoxy nucleotide method (Sanger et al., 1977 PNAS 74: 5463-5467) by using the T7 DNA polymerase (Sequenase US B CORP, Cleveland, OH) and α-^{35}S dATP (Amersham, England). The reaction products were loaded onto 6% denaturing polyacrylamide gels.

- Figure 10 : alignment of the amino acid sequences of VanC (SEQ ID NO:2), VanA (SEQ ID NO:4), Dd1A (SEQ ID NO:32) and Dd1B (SEQ ID NO:33). The identical (I) amino acids and the conservative (C) substitutions in the 4 sequences are indicated in the alignment. In order to classify the conservative substitutions, the amino acids were grouped as follows: RK, LFPMVI, STQNC, AGW, H, ED and Y. The regions of high homology corresponding to the domains 1, 2, 3 and 4 are underlined. The sequences corresponding to the peptides 1 and 2 are indicated by the arrows.

- Figure 11 : description of the oligonucleotides V1 (SEQ ID NO:9) and V2 (SEQ ID NO:10) (A) : Amino acid sequence of the peptides 1 (SEQ ID NO:36) and 2 (SEQ ID NO:37) of VanA and of the D-Ala-D-Ala ligases (SEQ ID NOS:36-39). The number of amino acids between the N-terminus and peptide 1, between the peptides 1 and 2 and the peptide 2 and the C-terminus is indicated. The identical amino acids between at least 2 of the 3 sequences are indicated in bold characters.

(B) : Target peptides (SEQ ID NOS:36-39) and deduced nucleotide sequence. X represents any base of the DNA. Peptide 2 in Dd1B (SEQ ID NO:39) differs from the target peptide at 2 positions (*).

(C) : Nucleotide sequence of V1 (SEQ ID NO:9) and V2 (SEQ ID NO:10). Alternate nucleotides and deoxyinosine (I) which may correspond to any base in the DNA, were used at the positions at which the nucleotide sequences coding for

the target peptides vary. The arrows indicate the direction of DNA synthesis. The oligonucleotides were synthesized by the methoxy-phosphoramidite method with a Biosystem DNA 380B machine (Applied Biosystem, Foster City, Ca). The DNA was isolated from bacterial lysates by extraction with hexadecyl trimethyl ammonium bromide (Inst. biotechnologies, Inc., New Haven, CO) (Le Bouguénec et al., 1990, J. Bacteriol. 172:727-734) and used as matrix for the amplification by means of PCR with a controlled heating system "Intelligent Heating Block" IBH101 (Hybarid Ltd., GB) according to the description of Mabilat et al. (1990, Plasmid 23:27-34). The amplification products were revealed by electrophoresis on a 0.8% gel, after staining with ethidium bromide.

- Figure 12: Inactivation by insertion of vanC. The vanC gene is shown by an open arrow and the internal EcoRI-HincII fragment of 690 bp is hatched. The DNA of pAT114 is shown by a thin line; the chromosomal DNA of PM4174 by a thick line; the arrows indicate the genes for resistance to the antibiotics: aphA-3 is the gene coding for the 3'-aminoglycoside phosphotransferase; erm is the gene coding for the ER^R methyl transferase.
(A) : The plasmid pAT217 was constructed by ligation of the EcoRI-HincII fragment of pAT216 to the suicide vector pAT114 (Trieu-Cuot et al., 1991, Gene 106:21-27), digested with EcoRI and SmaI.
(B) : vanC region of the chromosomal DNA of BM4174.
(C) : vanC region after integration of pAT217.

- Figure 13 : Southern blot analysis of the integration of pAT217 into the vanC gene of BM4174.
(left hand side) : Total DNA of BM4175 (line 2) and BM4174 (line 3) digested with EcoRI and resolved by means of electrophoresis on a 1% agarose gel. The DNA of the bacteriophage lambda digested with PstI was used as molecular mass standard (line 1). The DNA was transferred under vacuum to a Nytran membrane (Schleicher and Schül, Germany) by using a Trans-Vac TE80 apparatus (Höfer Scientific Instruments, San Francisco, CA) and bound to the membrane through the intermediary

of UV light. The hybridization was carried out with the probe C (Middle) or the probe aphA-3 specific for pAT114 (Lambert et al., 1985, Annales de l'Institut Pasteur/Microbiol. 136(b): 135-150).
(right hand side): the probes were labelled with ^{32}P by nick translation. The molecular masses (kb) are indicated.

- Figure 14 : alignment of the deduced amino acid sequences of VanS derived from E. faecium BM4147 (SEQ ID NO:40)(SEQ ID NO:41)(SEQ ID NO:42) and of PhoR and EnvZ from E.coli. The numbers on the left refer to the position of the first amino acid in the alignment. The numbers on the right refer to the position of the last amino acid of the corresponding line. The identical amino acids are placed in boxes. The dotted lines indicate gaps introduced in order to optimize their similarity. The dashes indicate the positions of the amino acid residues conserved in other HPK. The histidine residues in bold characters in section 1 are potential sites of auto-phosphorylation.

- Figure 15 : alignment of the deduced amino acid sequences of VanR from E. faecium BM4147, (SEQ ID NO:43) OmpR (SEQ ID NO:44) and PhoB (SEQ ID NO:45) from E. coli as well as that of CheY (SEQ ID NO:46) from Salmonella typhimurium. The numbers on the right indicate the position of the last amino acid of the corresponding line. The identical amino acids are placed in boxes. The dotted lines indicate the gaps introduced in order to optimize the homologies. The residues in bold characters correspond to the amino acids strongly conserved in the effector domains of other RR. The aspartic acid residue 57 of CheY is phosphorylated by the HPK associated with CheA.

I - IDENTIFICATION OF vanA

Materials and methods for the identification and characterization of the vanA gene

Bacterial strains and plasmids

The origin of the plasmids used is given in the table below.

Strain or plasmid	Source or reference
Escherichia coli	
JM83	Messing (1979)
AR1062	Rambach and Hogness (1977)
JM103	Hannshan (1983)
ST640	Lugtenberg and van Schijndel van-Dam (1973)
Enterococcus faecium	
BM4147	Leclercq et al (1988)
Plasmid pUC18	Norrander et al (1983)
pAT213	Brisson-Noel et al (1990)
pAT214	Described in this text

Preparation of the enterococcal membranes

Enterococcus faecium BM4147 was cultivated in 500 ml of heart-brain broth (BHI broth medium) until the optical density (OD_{600}) reached 0.7. Induction was effected with 10 µg/ml of vancomycin (Eli Lilly Indianapolis Ind). The subsequent steps were performed at 4°C. The cells were recovered by centrifugation for 10 minutes at 6000 g, washed with a TE buffer (0.01 M TRIS-HCl, 0.002 M EDTA, pH 7.0) and lysed by glass beads (100 µm in diameter) in a Braun apparatus for 2 minutes. The cell debris were separated by centrifugation for 10 minutes at 6000 g. The membranes were collected by centrifugation for 1 hour at

65000 g and resuspended in 0.5 ml of TE buffer.

Preparation of the minicells

Plasmids were introduced by transformation into the strain E. coli AR1062 prepared in the form of bacterial vesicles. The bacterial vesicles were recovered on sucrose gradients and the proteins were labelled with 50 µCi of [^{35}S]-L-methionine (Amersham, Great Britain) according to the method of Rambach and Hogness (1977, P.N.A.S. USA, 74; 5041-5045).

Preparation of the membrane fractions and the cytoplasmic fractions of E. coli

E. coli JM83 and strains derived from it were placed in culture in BHI medium until an optical density (OD_{600}) of 0.7 was attained, washed and suspended in a TE buffer. The cell suspension was treated by sonication (ultrasound) for 20 seconds at doses of 50 W in a cell fragmentation apparatus in a Branson B7 sonication apparatus and the intact cells were removed by centrifugation for 10 minutes at 6000 g. The supernatant was fractionated into membrane and cytoplasmic fractions by means of centrifugation for 1 hour at 100,000 g.

Electrophoresis on SDS-polyacrylamide gel (SDS-PAGE)

The proteins from the bacterial fractions were separated by means of SDS-PAGE on linear gradients of polyacrylamide gels (7.5% - 15%) (Laemmli 1970, Nature 227 : 680-685). The electrophoresis was carried out for 1 hour at 200 V, then for 3 hours at 350 V. The gels were stained with Coomassie blue. The proteins of the extracts were separated on 10% polyacrylamide gels and visualized by means of autoradiography.

Purification of the protein band and determination of the N-terminal sequence

The proteins of the membrane fractions of an induced culture of E. faecium BM4147 were separated by means of SDS-PAGE. The gel was electrotransferred for 1 hour at 200 mA to a polyvinylidene difluoride membrane (Immobilon Transfer, Millipore) by using a transfer apparatus (Electrophoresis Unit LKB 2117 Multiphor II) in accordance with the instructions of the manufacturer. The transferred proteins were stained with Ponceau red. The portion of membrane bearing the protein of interest was excised, centered on a Teflon filter and placed in the cartridge of a sequencer (Sequencer Applied Biosystems model 470A). The protein was sequenced by means of the automated Edman degradation (1967, Eur. J. Biochem. 1; 80-81).

Construction of plasmids

The plasmid pAT213 (Brisson-Noel et al., 1990, Antimicrob. Agents Chemother., 34; 924-927) consists of a EcoRI fragment of DNA of 4.0 kb of the enterococcal plasmid pIP816 cloned at the EcoRI site of a Gram-positive-Gram-negative shuttle vector pAT187 (Trieu-Cuot et al., 1987, FEMS Microbiol. Lett. 48; 289-294). In order to construct pAT214, the EcoRV-SacII DNA fragment of 1761 bp of pAT213 was purified, treated with the Klenow fragment of the DNA polymerase I of E. coli and ligated to the DNA of pUC18 which had previously been digested with SmaI and dephosphorylated (Figure 2). The cloning (Maniatis et al., 1982 Cold Spring Harbor Laboratory Press) was carried out with restriction endonucleases (Boehringer Mannheim and Pharmacia), with the T4 DNA ligase (Pharmacia) and alkaline phosphatase (Pharmacia) according to the instructions of the manufacturer.

Subcloning in M13 and nucleotide sequence

The DNA restriction fragments were subcloned in the polylinker of the replicative forms of the derivatives mp18 and mp19 of the bacteriophage M13 (Norrander et al., 1983, Gene 26; 101-106), obtained from Pharmacia P-L Biochemicals. E.coli JM103 was transfected with recombinant phages and the single-stranded DNA was prepared. The nucleotide sequencing was carried out by the enzymatic di-deoxy nucleotide method (Sanger et al., 1977, P.N.A.S. USA 74; 5463-5467) by using a T7 DNA polymerase (Sequenase, United States Biochemical Corporation, Cleveland, Ohio) and [α-^{35}S]dATP (Amersham, Great Britain). The reaction products were revealed on 6% polyacrylamide gels containing a denaturing buffer.

Data-processing analysis and data on the sequence

The complete DNA sequence was assembled by using the computer programs DBCOMP and DBUTIL (Staden, 1980, Nucleic Acids Res 8; 3673-3694). The protein data bank PSEQIP of the Pasteur Institute was screened using an algorithm developed by Claverie (1984, Nucleic Acids Res 12; 397-407). The alignments between the pairs of amino acid sequences were constructed using the algorithm of Wilbur et al (1983, P.N.A.S. USA 80; 726-730). The statistical significance of the homology was evaluated with the algorithm of Lipman and Pearson (1985, Science 227; 1435-1440).

For each comparison 20 amino acid sequences were used to calculate the mean values and the standard deviations of the random results.

Genetic complementation tests

The plasmids were introduced by transformation into E.coli ST640, a temperature-sensitive mutant with an unmodified D-ala-D-ala ligase (Lugtenberg et al 1973, J. Bacteriol 110; 26-34). The transformants were selected at 30°C on plates containing 100 µg/ml of ampicillin and the presence of the plasmid DNA of the expected size and the restriction maps were verified. Single colonies grown at 30°C in BHI broth medium containing ampicillin were placed on a BHI agar medium containing both 100 µg/ml of ampicillin and 50 µM of isopropyl-1-thio-ß-D-galacto-pyranoside (IPTG) and the plates were incubated at a permissive temperature of 30°C and at a non-permissive temperature of 42°C. The complementation test was considered to be positive if the colonies were present after 18 hours of incubation at 42°C.

RESULTS

Identification of the VanA protein and its N-terminal sequence

The membrane fractions of the E. faecium BM4147 cells placed in culture, on the one hand, under conditions of induction, and, on the other, in the absence of induction, were analysed by means of SDS-PAGE. The sole difference which could be detected, related to the exposure to sub-inhibitory concentrations of vancomycin, was the marked intensification of a band which corresponded to a protein of an estimated molecular weight of about 40 kDa. In the induced cells and in the non-induced cells, the protein band represents the same protein because this band is absent from membranes of a derivative of BM4147 which has lost the pIP816 plasmid. The inducible protein, designated as VanA, was purified after SDS-PAGE and automated Edman degradation was carried out on a 50 pmol. sample. Nine amino acids of the N-terminal sequence of VanA were identified: Met Asn Arg Ile Lys Val Ala Ile Leu. (SEQ ID NO:47)

Sub-cloning of the vanA gene

The insert of 4.0 kb of the plasmid pAT213 bears the determinant for resistance to the glycopeptides of E. faecium BM4147. Various restriction fragments of this insert were subcloned in pUC18 and the recombinant plasmids specific for vanA in E. coli were identified by SDS-PAGE analysis of the proteins of the cytoplasmic and membrane fractions or of the extracts of the bacterial vesicles. This approach was used since E. coli is intrinsically resistant to the glycopeptide. The EcoRV-SacII insert of the pAT214 plasmid (Figure 2) codes for a unique polypeptide of 40 kDa which migrates together with VanA, derived from the membrane preparations of E. faecium BM4147.

Nucleotide sequence of the insert in pAT214 and identification of the vanA coding sequence

The nucleotide sequence of the EcoRV-SacII insert of 1761 bp in pAT214 was determined on both strands of the DNA according to the strategy described in Figure 2. The location of the termination codons (TGA, TAA, TAG) in three reading frames on each DNA strand showed the presence of a unique open reading frame (ORF) which was sufficiently long to code for the VanA protein. This reading frame ORF is located between the TAA codon at position 281 and the TAG codon at position 1406. The amino acid sequence deduced for ORF was compared with that of the N-terminus of VanA. The nine amino acids identified by protein sequencing are encoded in the nucleotide sequence beginning with the ATG (methionine) codon at position 377 (Figure 3). This codon for the initiation of translation is preceded by a sequence (TGAAAGGAGA) (SEQ ID NO:48) characteristic of a ribosomal binding site (RBS) in Gram-positive bacteria which is complementary to the 8 bases of the rRNA of the 16S subunit of Bacillus subtilis in its sequence (3'OH UCUUUCCUCC 5') (SEQ ID NO:49) (Moran et al., 1982, Mol. Gen. Genet. 186; 339-346). In this ORF, there is no other ATG or GTG initiation codon between the positions 281 and 377. The sequence of 1029 bp which extends from the ATG codon at position 377 to the TGA codon at position 1406 codes for a protein

containing 343 amino acid residues. The calculated molecular weight of this protein is 37400 Da, which is in agreement with the estimation of 40 kDa obtained by SDS-PAGE analysis.

Homology of the amino acid sequences of VanA and the D-ala-D-ala ligase enzymes

The screening of the protein data bank PSEQIP has shown the existence of a sequence homology between VanA and the D-ala-D-ala ligases of E.coli (ECOALA, Robinson et al., 1986, J. Bacteriol. 167; 809-817) and of Salmonella typhimurium (DALIG, Daub et al., 1988, Biochemistry 27; 3701-3708). The calculated percentage of homology between pairs of proteins was included between 28% and 36% for the identical amino acids and between 48% and 55% by taking into consideration homologous amino acids. VanA (SEQ ID NO:4) and DALIG are more closely related. The statistical significance of these similarities wa evaluated by aligning VANA and sequences containing the same composition of amino acids as DALIG or ECOALA (Lipman and Pearson, 1985, Science 227; 1435-1440).

Genetic complementation test for the activity of D-ala-D-ala ligase

The E.coli strain ST640 is a thermosensitive mutant exhibiting a deficient D-ala-D-ala ligase activity (Lugtenberg et al., 1973, J. Bacteriol. 113: 96-104). The plasmids pUC18 and pAT214 were introduced into E.coli ST640 by transformation. The strains ST640 and ST640 (pUC18) grew normally only at the permissive temperature (30°C) whereas E.coli ST640 (pAT214) grew both at the permissive temperature and at the non-permissive temperature (42°C).

This test shows that VANA is functionally related to the D-Ala-D-Ala ligases in E.coli and is probably capable of catalysing

the same ligation reaction as DALIG.

II - VanS-VanR two-component regulation system for the control of the synthesis of depsipeptides of the precursor of peptidoglycans

MATERIALS AND METHODS

Strains, plasmids and conditions of culture

The restriction fragments of pIP816 (Tra^-, Mob^+, Vm^r) were cloned in derivatives of the vector pAT29 which constitutes a shuttle vector between the Gram-positive and Gram-negative bacteria (oriR pAMß1, oriR pUC, oriT RK2, spc, lacZ) (Trieu-Cuot et al., 1990, Nucleic Acids Res. 18:4296). This vector was constructed by the inventors and used to transform the strain E.coli JM103 ((lac-proAB), supE, thi, strA, sbcB15, endA, hspR4, F traD36, proAB, $LacI^q$, lacZ M15) (Messing et al., 1983, Methods Enzymol. 101:20-78). The plasmid DNA was prepared by an alkaline lysis protocol on a small scale (Sambrook et al., 1982, Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor NY) and introduced by electroporation (Cruz-Rodz A.L. et al., 1990, Mol. Gen. Genet. 224: 152-154) in E.faecalis JH2-2 (Fus^R, Rif^R) (Jacob A.E. et al., 1974, J. Bacteriol. 117: 360-372), by using a Gene Pulser apparatus (Bio-Rad Laboratories, Richmond, California). The restriction profiles of the purified plasmids from E. faecalis and E. coli were compared in order to detect possible rearrangements of DNA.

The integrative plasmid pAT113 (Mob^+, Em^R, Km^R, oriR PACYC184, attTn1545, LacZ) (Trieu-Cuot et al., Gene 106: 21-27) carries the joined ends of the transposon Tn1545. This vector does not replicate in Gram-positive bacteria but is integrated into the chromosome of the host by illegitimate recombination mediated by the integrase of

Tn1545 or of Tn916 (Trieu-Cuot et al. previously mentioned). The integrative plasmids were introduced into E. faecalis BM4148 (strain JH2-2::Tn916) by means of electroporation. This strain is modified by the transposon Tn917 described by Franque A.E. et al. (1981, J. Bacteriol. 145: 494-502).

The cultures were grown in brain-heart broth (BHI - Brain Heart Infusion Broth) or on agar at 37°C. The method of Steers et al (Antibiot. Chemother. Basel. 9: 307-311) was used to determine the minimal inhibitory concentrations (MICs) of the antibiotics on a Mueller-Hinton gelose agar medium.

Recombinant DNA procedures

The cleavage of DNA with restriction endonucleases (Boehringer Mannheim and Pharmacia), the purification of the DNA restriction fragments from agarose gels, the conversion of the cohesive ends to blunt ends with the Klenow fragment of the DNA polymerase I of E.coli (Boehringer Mannheim), the dephosphorylation of the ends of the DNA with calf intestinal phosphatase (Boehringer Mannheim), the ligation of the DNA fragments with the T4 DNA ligase (Amersham) were carried out according to the standard methods of Sambrook et al (1982, Molecular Cloning, a Laboratory Manual. Cold Spring Harbor Laboratory. Cold Spring Harbor NY).

Construction of plasmids

The origin of the vectors and the inserts used for the recombinant plasmids constructed here is the following:

(i) vector pAT78 for the recognition of the promoter: the amplified DNA of the cat gene for chloramphenicol acetyltransferase of the plasmid pC194 of Staphylococcus

aureus (Horinouchi et al., 1982, J. Bacteriol. 150: 815-825) was inserted between the PstI and SphI restriction sites of the shuttle vector pAT29. Amplification by means of the polymerase chain reaction was carried out by means of primers A1 and A2 which were synthesized by the methoxy phoshoramidite method (Mabilat et al., 1990, Plasmid 23: 27-34). The sequence of the primer A1 (SEQ ID NO: 50) (5' GCTGCAGATAAAAATTTAGGAGG) is composed of a PstI recognition site (underlined) and 18 bases (positions 6 to 23) of pC194 which include the ribosomal binding site (RBS ; AGGAGG positions 18 to 23) of the cat gene. The sequence of the primer A2 (SEQ ID NO: 51) (5' CGCATGCTATTATAAAA GCCAGTC) contains the SphI cleavage site (underlined) and is complementary (positions 8 to 24) to 17 bases at the 3' end of the cat gene. The triplet ATT at positions 9 to 11 correponds to the TAA stop codon of cat. The DNA fragments amplified with the primers A1 and A2 hence consist of an open reading frame (orf) and a ribosomal binding site for CAT (positions 1234 to 1912 according to the numbering of Horinouchi et al. (1982, J. Bacteriol. 150: 815-825) flanked by the PstI and SphI sites. The position 1234 is located at the interior of the loop of the secondary structure of the mRNA which blocks translation in the absence of chloramphenicol. Thus, the amplified sequence does not contain the cat promoter nor the sequence complementary to the RBS which is essential for the regulation of translation Ambulos, N.P. et al., 1984, Gene 28: 171-176).

(ii) expression vector pAT79: the ClaI-BssHII fragment of 243 bp bearing the P2 promoter of the aphA-3 gene of the enterococcal plasmid pJH1 (Caillaud et al., 1987, Mol. Gen. Genet. 207: 509-513) was inserted between the EcoRI and SacI restriction sites of pAT78.

(iii) plasmid pAT80 and its derivatives: the BglII-XbaI fragment of 5.5 kb of pIP816 was inserted between the BamHI

and XbaI sites of pAT78. The resulting plasmid, designated as pAT80 was partially digested with HincII and ligated with the EcoRV fragment containing a gene related to the apha-I gene of the transposon Tn903 (Oka A. et al., 1981, J. Mol. Biol. 147:217-226. This fragment contains the aphA-I gene which codes for the 3'aminoglycoside phosphotransferase of type I conferring resistance to kanamycin. The insertion of aphAI was carried out at three different sites in pAT80, generating the plasmids pAT81, pAT83 and pAT85. The cassettes BamHI and EcoRI containing aphA-I were inserted at the BamHI (to form the plasmid pAT84) and EcoRI (to form the plasmid pAT82) sites of pAT80.

(iv) plasmids pAT86, pAT87, pAT88 and pAT89: the plasmid pAT86 was constructed by cloning the EcoRI-SacII fragment of 2.803 bp of pAT80 coding for VanH and VanA at a SmaI site of pAT79. pAT87 was obtained by inserting the EcoRI-XbaI fragment of 3.4 kb of pAT80 upstream from the cat gene of the detection vector of promoter pAT78. The plasmid pAT88 resulted from the ligation of pAT78 digested with EcoRI and BamHI to the EcoRI-BamHI fragment of 1.731 bp of pAT80. The BglII-AccI fragment (positions 1 to 2356) of pAT80 was inserted into the polylinker of the integrative vector pAT113, generating pAT89.

Sub-cloning in M13 and sequencing

The DNA restriction fragments were subcloned in a polylinker of replicative derivatives of the bacteriophage M13, these derivatives being called mp18 and mp19 (Norrander et al., 1983, Gene 26:101-106). E.coli JM103 was transfected with the recombinant phages and a single-stranded DNA was prepared. The sequencing of the nucleotides was carried out according to the conditions described by Sanger et al. (Proc. Natl.

Acad. Sci. USA, 1977, 74: 5463-5467) by using the modified T7 DNA polymerase (Sequenase, United States, Biochemical Corporation Cleveland OH) and α-^{35}S/dATP (Amersham). The reaction products were resolved on gradient gels of polyacrylamide in a 6% buffer.

Enzymatic test

The JH2-2 derivatives of E. faecalis were grown to an optical density OD_{600} of 0.7 in a BHI broth supplemented with spectinomycin (300 µg/ml). The cells were treated with lysozyme, lysed by sonication and the cell debris were centrifuged for 45 minutes at 100,000 g according to the description given by Courvalin et al. (1978, Antimicrob. Agents Chemother. 13:716-725). The formation of 5-thio-2-nitrobenzoate was measured at 37°C in the presence and in the absence of chloramphenicol and the specific CAT activity was expressed in micromole per minute and per milligram of proteins (Shaw et al., 1975, Methods Enzymol. 43:737-755).

RESULTS

The vanH and vanA genes of pIP816 were cloned in a plasmid pAT79 under the control of the heterologous promoter P2 (Caillaud et al., 1987, Mol. Gen. Genet. 207:509-513) and the plasmid pAT86 formed did not confer resistance to vancomycin on the strain E. faecalis JH2-2. These genes are thus not sufficient for the synthesis of peptoglycan in the absence of the antibiotic. Different restriction fragments of pIP816 were cloned in the vector pAT78. The BglII-XbaI fragment of 5.5 kb of pAT80 is the smallest fragment obtained which conferred resistance to vancomycin.

Nucleotide sequence of the vanR and vanS genes

The sequence of the insert in pAT80 was determined on both strands of the DNA from the BglII site to the ATG initiation codon for the translation of VanH. Two open reading frames (orf) were detected within the sequence of 2475 bp: the first open reading frame extends from the nucleotide 386 to the nucleotide 1123; at position 431 a sequence characteristic of the RBS sequences in Gram-positive bacteria is found, 6 base pairs upstream from the ATG initiation codon for translation (TGAAAGGGTG) (SEQ ID NO:52); the other initiation codons for translation in this orf are not preceded by this type of sequence. The sequence of 693 bp extending from the ATG codon at position 431 to the TAA codon at position 1124 is capable of coding for a protein of 231 amino acids with a molecular mass of 26,612 Da which is designated as VanR.

In the case of the second open reading frame (from nucleotide 1089 to nucleotide 2255) the amino acid sequence deduced from the first initiation codon in phase (TTG at position 1104) would code for a protein of 384 amino acids having a molecular mass of 43,847 Da and designated as VanS. The TTG codon at position 1116 and the ATG codon at position 1164 are in-phase initiation codons for translation preceded by sequences with low complementarity with the 3'OH terminus of the 16S sub-unit of the rRNA of B. subtilis (GGGGGGTTGG-N8-TTG (SEQ ID NO:53) and AGAACGAAAA-N6-ATG (SEQ ID NO:54), respectively).

Between the last codon of vanS and the initiation codon ATG for the translation of vanH a sequence of 217 bp is to be observed which contains a repeated reverse sequence of 17 bp. This sequence does not function as a terminator of strong transcription.

The comparison of the sequences obtained with data bases has shown that the conserved amino acid residues identified by Stock et al. (1989, Microbiol. Rev. 53:450-490) in the kinase domain of 16 HPK (Histidine Protein Kinase) were detected in the C-terminal part of VanS. VanS (SEQ ID NO:14) possesses two groups of hydrophobic amino acids in the

N-terminal region. The histidine residue 164 of VAnS is aligned with the residue His216 of PhoR (SEQ ID NO:41) (Makino et al., 1986, J. Mol. Biol. 192: 549-556) and His 243 of EnvZ (SEQ ID NO:42) (Comeau et al., 1985, 164:578-584) which are presumed sites of autophosphorylation in these proteins.

Similarly, the amino acids 1 to 122 of VanR (SEQ ID NO:12) exhibit similarities with the effector domains of response regulators RR. The aspartic acid 53 of VanR might be a phosphorylation site because this residue is aligned with Asp 57 of Che Y (SEQ ID NO:46) which is phosphorylated by HPK associated with CheA and corresponds to an invariant position in other proteins of the RR type (Stock et al previously mentioned). VanR might belong to the sub-class OmpR-PhoB of RR which activates the initiation of transcription mediated by the RNA polymerase containing the 70S factor of E.coli (Stock et al. previously mentioned).

Inactivation of the van genes by insertion

Cassettes of resistance to kanamycin inserted in the group of van genes in the plasmid pAT80 have shown the following: the insertion in vanR suppresses resistance to vancomycin and chloramphenicol; VanR is an activator of transcription necessary for the expression of the genes for resistance to vancomycin. The inactivation of vanS leads to a two-fold reduction of the minimal inhibitory concentration (MIC) of chloramphenicol and to a three-fold reduction of the specific CAT activity but the minimal inhibitory concentration of vancomycin remains unchanged. Hence, VanS is necessary to produce a high level of transcription of the genes for resistance to vancomycin although it is not required for the expression of the phenotype of resistance to vancomycin.

Derivatives of pAT80 bearing insertions in vanH (pAT83), vanA (pAT84) or in the region 1.0 kb downstream from vanA (pAT85) have made it possible to obtain resistance to chloramphenicol but not to vancomycin. This dissociated phenotype corresponds to the inactivation

of genes coding for enzymes which synthesize the depsipeptide precursors necessary for the assembly of the bacterial cell walls in the presence of vancomycin.

Downstream from the vanA gene the presence of an inactivated orf has been detected in pAT85 in the region of the sequence of 365 bp after the TGA codon of vanA and before the SacII site and this orf contains an in-phase ATG initiation codon preceded by a RBS-like sequence. This sequence codes for a protein necessary for resistance to the glycopeptide, designated as VanX and which comprises maximally about 330 amino acids.

Trans-activation of the transcription of the van genes

The integrative plasmid pAT89 coding for VanR and VanS was introduced into the chromosome of *E. faecalis* BM4138. The plasmid pAT87 bearing the genes vanH, vanA and vanX cloned upstream from the cat gene lacking the promoter for pAT78 conferred resistance to vancomycin on this strain but not to *E. faecalis* JH2-2. The level of expression of the cat gene of pAT87 in the strains BM4138::pAT89 and JH2-2 indicated that VanR activates the transcription of the reporter gene localized at the 3' end of the group of van genes. Similar levels of CAT synthesis were observed for pAT88 which bears a transcription fusion between the 5' parts of vanA and the cat gene. These results show that in *E. faecalis* BM4138::pAT89 (pAT87) VanR and VanS encoded in the chromosome activate in a trans manner the transcription of vanA, vanH and vanX of pAT87 making possible the production of resistance to vancomycin.

Moreover, it has been observed that the expression of the gene was essentially constitutive when vanR and vanS were borne by a multicopy plasmid pAT80 and weakly inducible by vancomycin when the genes for the regulatory proteins were present on the chromosome of the host.

III - Characterization of the sequence of the vanC gene of Enterococcus gallinarum BM4174

Definition and use of universal primers for the amplification of genes coding for D-Ala-D-Ala ligases and related proteins implicated in resistance to vancomycin

The protein VanA necessary for the expression of a high level of resistance to the glycopeptides in E. faecium BM4147 shares a similarity of about 28 to 36% as regards its amino acids with the D-Ala-D-Ala ligases of E.coli but possesses a different substrate specificity from that of these ligases. Peptides designated as 1 and 2 which are conserved in the sequences of the DdlA and DdlB ligases (Zawadzke, 1991 Biochemistry 30:1673-1682) of E.coli and in the protein VanA were selected in order to synthesize universal primers intended to amplify internal fragments of genes coding for D-Ala-D-Ala ligases or related enzymes. The peptide targets GEDG(S/T) (I/L)QG and NT(I/L)PGFT were translated back as is shown in Figure IV.1 in order to obtain degenerate oligonucleotides V1 and V2. As the peptides 1 and 2 of VanA, DdlA and DdlB are separated by amino acid sequences of similar length, the predicted size for the amplification product was about 640 bp.

Amplification by means of PCR with the DNA of E.coli JM83 and of E. faecium BM4147 made it possible to amplify products corresponding to the expected size which have then been purified and cloned in the bacteriophage M13mp10 (Norrander et al., 1983, Gene 26:101-106). The sequencing of the insert obtained with E.coli JM83 has shown that the product of PCR was an internal fragment of ddlA. A probe generated starting from a recombinant phage obtained with the amplification fragment of BM4147 was used for the Southern blot analysis of a DNA of BM4147 and BM4147-1 which is a derivative of BM4147 sensitive to vancomycin and which lacks the plasmid pIP816 (Leclercq et al., 1988, N. Engl. J. Med. 319:157-161). The probe hybridized with the EcoRI DNA fragment of 4 kb from BM4147 but not with the DNA from E. faecium BM4147-1. As the vanA gene is borne by the EcoRI fragment

of 4 kb from pIP816, these results indicate that the primers also make possible the amplification of a part of vanA. Thus the oligonucleotides V1 and V2 may amplify fragments of genes coding for different proteins related to the D-Ala-D-Ala ligases, and may do this in different species.

Amplification, cloning and sequencing of the vanC gene

Amplification by means of PCR was carried out on the total DNA of E. gallinarum BM4174 and the amplification product obtained of about 640 bp was cloned in the bacteriophage M13mp10. The single-stranded DNA isolated from the recombinant phage was used to construct a probe C (Hu et al., 1982, Gene 17:2171-2177). In Southern analysis the probe hybridized with a PstI fragment of 1.7 kb from BM4174 but not with the DNA of BM4147 and BM4147-1.

The DNA of BM4174 was digested with PstI and fragments of 1.5 and 2 kb were purified by electrophoresis on agarose gel and cloned in pUC18 (Norrander et al., 1983, mentioned previously). The recombinant plasmids were introduced into E.coli JM83 by transformation and screened by hybridization on colonies (Sambrook et al., 1989, Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY) by using the probe C. A homology was detected with a transformant harbouring a plasmid called pAT216 which contained a PstI insert of 1.7 kb. The sequence of the SacI-PstI part of 1347 bp of the insert of pAT216 was determined on both strands of the DNA. The location of the termination codons in the three reading frames of each strand of DNA revealed the presence of an ORF phase located between the TGA codons at positions 47 and 1244. The initiation codon of transcription ATG at position 215 is preceded by a sequence GAAAGGAAGA characteristic of the RBS sequences complementary to the RNA of the 16S subunit of B. subtilis (Moran et al., 1982, Mol. Gen. Genet. 186:339-346). The sequence of 1029 bp which extends from the ATG codon at position 215 to the TGA codon at position 1244 might code for a

protein of 343 amino acids having a calculated molecular mass of 37504 Da designated as VanC. A sequence homology was detected between VanC, VanA and the D-Ala-D-Ala ligases of E.coli. In particular, four domains of strong homology previously found between VanA and the D-Ala-D-Ala ligases of the enterobacteria are also present in VanC. The percentage of identical amino acids calculated for these proteins taken two at a time varied between 29 and 38%. The alignment of the four sequences revealed the presence of 57 invariant amino acids which include the conserved residues of the peptides 1 and 2 used to define the oligonucleotide probes V1 and V2.

Inactivation of the vanC gene by insertion

In order to evaluate the contribution of vanC to resistance to vancomycin in E. gallinarum BM4174, the vanC gene was inactivated by insertion. A EcoRI-HincII fragment of 690 bp, internal to vanC was cloned in pAT114 which does not replicate in Gram-positive bacteria. The resulting pAT217 plasmid was introduced into BM4174 by electroporation (Cruz-Rodz et al., 1990, Mol. Gen. Genet. 224:152-154) and the clones supposed to result from a homologous recombination leading to the integration of pAT217 into vanC were selected on erythromycin. The clone BM4175 was compared with BM4174 by Southern hybridization using the probe C and aphA-3 specific for pAT114. The two probes hybridized with the EcoRI fragment of 8.6 kb from BM4175. The probe C hybridized with a fragment of 2.5 kb from BM4174 whereas no signal was observed with the probe aphA-3. The results indicate that the plasmid pAT217 of 6.1 kb was integrated into the vanC gene. The determination of the minimal inhibitory concentration of vancomycin for BM4174 (16 mg/l) and BM4175 (2 mg/l) indicated that the inactivation by insertion in vanC abolishes resistance to vancomycin.

VanC is thus required for resistance to vancomycin. It may thus be supposed that this protein synthesizes a dipeptide or a depsipeptide which is incorporated into the precursors of peptido-

glycans and is not recognized by vancomycin.

The sequences which are the object of the invention are given in the following pages after the list of the sequences containing the description of these sequences. In this list of the sequences, the proteins are identified with respect to the position of the nucleotide bases corresponding to the amino acids of the extremities of the proteins.

List of the sequences

(contained in the sequences I (Ia, Ib), II presented below or in the sequence shown in Figure 5).

Amino acid sequences

SEQ ID NO 2 (VanH) : sequence of the first resistance protein, corresponding to the amino acid sequence of the open reading frame No. 3, starting at the base 3501 and terminating at the base 4529, containing the sequence coding for the vanH gene between the bases 3564 and 4529 with respect to the sequence shown in Figure 5 or corresponding to the sequence between the positions of the nucleotides 6018 and 6983 of the sequence Ia.

SEQ ID NO 4 (VanA) : sequence of the VanA protein, corresponding to the amino acid sequence of the open reading frame No. 1, starting at the base 4429 and terminating at the base 5553 with respect to the sequence shown in Figure 5 or corresponding to the sequence between the positions of the nucleotides 6977 and 7807 of the sequence Ia.

SEQ ID NO 6 (VanX) : sequence of the third resistance protein, corresponding to the amino acid sequence of the open reading frame No. 3, starting at the base 5526 and terminating at the base 6167 with respect to the sequence shown in Figure 5 or corresponding to the sequence between the positions of the nucleotides 7816 and 8621 of the sequence Ia.

SEQ ID NO 12 (VanR) : sequence of the regulatory protein R, corresponding to the amino acid sequence of the open reading frame No. 1, starting at the base 1477 and terminating at the base 2214 with respect to the sequence shown in Figure 5 or corresponding to the sequence between the positions of the nucleotides 3976 and 4668 of the sequence Ia.

SEQ ID NO 14 (VanS) : sequence of the sensor protein S, corresponding to the amino acid sequence of the open reading frame No. 2, starting at the base 2180 and terminating at the base 3346 with respect to the sequence shown in Figure 5 or corresponding to the sequence between the positions of the nucleotides 4648 and 5800 of the sequence Ia.

SEQ ID NO 19 : sequence of the transposase corresponding to the amino acids included between the nucleotides 150 and 3112 of the sequence Ib.

SEQ ID NO 21 : sequence of the resolvase comprising the amino acids situated between the positions of the nucleotides 3187 and 3759 of the sequence Ia.

SEQ ID NO 23 : VanY sequence comprising the amino acids situated between the positions of the nucleotides 9046 and 9960 of the sequence Ia.

SEQ ID NO 25 : VanZ sequence comprising the amino acids situated between the positions of the nucleotides 10116 and 10598 of the sequence Ia.

SEQ ID NO 8 : VanC amino acid sequence shown in list II.

- Nucleotide sequences

SEQ ID NO 15 : nucleotide sequence containing the sequence coding for the 5 proteins as well as the flanking sequences, shown in Figure 5.

SEQ ID NO 15 : sequence containing the sequence coding for the 3 resistance proteins as well as the flanking sequences and starting at the base 3501 and terminating at the base 6167, shown in Figure 5.

SEQ ID NO 1 : sequence of the vanA gene, starting at the base 4429 and terminating at the base 5553 of the sequence shown in Figure 5, or corresponding to the nucleotide sequence situated between the

nucleotides 6977 and 7807 of the sequence Ia.

SEQ ID NO 9 : sequence coding for the first resistance protein called VanH, starting at the base 3501 and terminating at the base 4529, in particular the sequence vanH, the coding sequence of which is located between the bases 3564 and 4529 of the sequence shown in Figure 5, or corresponding to the nucleotide sequence situated between the nucleotides 6018 and 6983 of the sequence Ia.

SEQ ID NO 10 : sequence coding for the third resistance protein VanX, starting at the base 5526 and terminating at the base 6167 of the sequence shown in Figure 5, or corresponding to the nucleotide sequence situated between the nucleotides 7816 and 8621 of the sequence Ia.

SEQ ID NO 11 : sequence of the transposon coding for the transposase, the resolvase, vanR, VAnS, VanH, VanA, VanX, VanY and VanZ and containing the repeated reverse sequence of 38 bp at its N- and C-termini and corresponding to the sequence Ia.

SEQ ID NO 12 : sequence coding for the transposase, starting at the base 150 and terminating at the base 3112 of the sequence Ib.

SEQ ID NO 13 : sequence coding for the resolvase, starting at the base 3187 and terminating at the base 3759 of the sequence Ia.

SEQ ID NO 14 : sequence coding for VanY, starting at the base 9046 and terminating at the base 9960 of the sequence Ia.

SEQ ID NO 15 : sequence coding for VanZ, starting at the base 10116 and terminating at the base 10598 of the sequence Ia.

SEQ ID NO 21 : sequence coding for VanC, shown in the list II in relation to the protein VanC.

SEQ ID NO 16 : complete sequence Ia of the transposon of E. faecium, starting at the base 1 and terminating at the base 10851.

SEQ ID NO 11 : sequence coding for the protein VanR, starting at the base 3976 and terminating at the base 4668 of the sequence Ia.

SEQ ID NO 13 : sequence coding for the protein VanS, starting at the base 4648 and terminating at the base 5800 of the sequence Ia.

I. Nucleotide sequence of the transposon and translation

Ia. (+) Strand

1
GGG GTA GCG TCA GGA AAA TGC GGA TTT ACA ACG CTA AGC CTA TTT TCC TGA CGA ATC CCT
61
CGT TTT TAA CAA CGT TAA GAA AGT TTT AGT GGT CTT AAA GAA TTT AAT GAG ACT ACT TTC
121
TCT GAG TTA AAA TGG TAT TCT CCT AGT AAA TTA ATA TGT TCC CAA CCT AAG GGC GAC ATA
181
TGG TGT AAC AAA TCT TCA TTA AAG CTA CCT GTC CGT TTT TTA TAT TCA ACT GCT GTT GTT
241
AGG TGG AGA GTA TTC CAA ATA CTT ATA GCA TTG ATA ATT ATG TTT AAA GCA CTG GCT CTT
301
TGC AAT TGA TGC TGT ATG GTG CGT TCT CTA AGC TCA CCT TGT TTT CCG AAG AAA ATA GCT
361
CTT GCC AAT CCA TTC ATG GCT TCT CCT TTA TTC AAT CCT CTT TGT ATT TTT CTT CTT AAT
421
GAT TCA TCC GAT ATA TAA TTC AAA ATA AAG ATC GTT TTT TCT ATT CGG CCC ATC TCA CGT
481
AAG GCT GTA GCT AAG CTG TTT TGT CTT GAA TAG GAA CCT AGC TTC CCC ATA ATA AGG GAT
541
GCT GAA ACT GTT CCC TCC CTT ATA GAA TGA GCT AAT CGC AAA ACA TCC TCA TAA TTT TCT
601
TTA ATG ACC TTT GTA TTT ATT TGT CCA CGT AAA ATG GCT TCT AGT TTT GGA TAC TCA CTT

661
GCT TTA TCT ATC GTA AAT AAT TTT GAG TCC GAT AAA TCC CTT ATT CTT GGG GCA AAT TTA
721
AAT CCT AAT AAA TGA GTC AGT CCG AAT ATT TGG TCA GTG TAA CCG GCA GTG TCT GTA TAA
781
TGT TCC TCT ATG TTT AGA TCC GTC TCA TGA TGT AAC AAA CCA TCC AAA ACA TGA ATC GCA
841
TCT CTT GAA TTA GTA TGA ATA ATC TTT GTG TAG TAA GAA GAG AAT TGA TCA CTT GTA AAT
901
CGG TAG ATG GTG GCT CCT TTT CCA GTT CCA TAA TGT GGA TTT GCA TCT GCA TGT AGT GAT
961
GAA ACA CCT AGC TGC ATT CTC ATA CCA TCT GAC GAA GAT GTT GTA CCG TCG CCC CAA TAG
1021
AAA GGC AAT TGT AAT TTA TGA TGA AAG TTT ACT AAT ATG GCT TGG GCT TTA TTC ATG GCA
1081
TCT TCA TAC ATG CGC CAT TGA GAT ACA TTG GCT AGT TGC TTA TAT GTA AGT CCG GGT GTG
1141
GCT TCG GCC ATC TTG CTC AAG CCA ATA TTC ATT CCC ATT CCT AAA AGG GCA GCC ATG ATA
1201
ATG ATT GTT TCT TCC TTA TCT GGT TTT CGA TTA TTG GAA GCA TGA GTG AAT TGC TCA TGA
1261
AAT CCT GTT ATA TGG GCC ACA TCC ATG AGT AAA TCA GTT AAT TTT ATT CTT GGT AGC ATC
1321
TGA TAA AGG CTT GCA CTA AAT TTT TTT GCT TCT TCT GGA ACA TCT TTT TCT AAG CGT GCA
1381
AGT GAT AGC TTT CCT TTT TCA AGA GAA ACC CCA TCT AAC TTA TTG GAA TTG GCA GCT AAC
1441
CAC TTT AAC CTT TCA TTA AAG CTG CTG GTT CTC TCC GTT ATA TAA TCT TCG AAT GAT AAA

1501
CTA ACT GAT AAT CTC GTA TTC CCC TTC GAT TGA TTC CAT GTA TCT TCC GAA AAC AAA TAT
1561
TCC TCA AAA TCC CTA TAT TGT CTG CTG CCA ACA ATG GAA ACA TCT CCT GCC CGA ACA TGC
1621
TCC CGA AGT TCT GTT AAA ACA GCC ATT TCA TAG TAA TGA CGA TTA ATT GTT GTA CCA TCA
1681
TCC TCG TAT AAA TGT CTT TTC CAT CGT TTT GAA ATA AAA TCC ACA GGT GAG TCA TCA GGC
1741
ACT TTT CGC TTT CCA GAT TCG TTC ATT CCT CGG ATA ATC TCA ACA GCT TGT AAA AGT GGC
1801
TCA TTT GCC TTT GTA GAA TGA AAT TCC AAT ACT CTT AAT AGC GTT GGC GTA TAT TTT CTT
1861
AGT GAA TAA AAC CGT TTT TGC AGT AAG TCT AAA TAA TCA TAG TCG GCA GGA CGT GCA AGT
1921
TCC TGA GCC TCT TCT ACT GAA GAG ACA AAG GTA TTC CAT TCA ATA ACC GAT TCT AAA ACC
1981
TTA AAA ACG TCT AAT TTT TCC TCT CTT GCT TTA ATT AAT GCT TGT CCG ATG TTC GTA AAG
2041
TGT ATA ACT TTC TCA TTT AGC TTT TTA CCG TTT TGT TTC TGG ATT TCC TCT TGA GCC TTA
2101
CGA CCT TTT GAT AAC AAA CTA AGT ATT TGC CTA TCA TGA ATT TCA AAC GCT TTA TCC GTT
2161
AGC TCC TGA GTA AGT TGT AAT AAA TAG ATG GTT AAT ATC GAA TAA CGT TTA TTT TCT TGA
2221
AAG TCA CGG AAT GCA TAC GGC TCG TAT CTT GAG CCT AAG CGA GAC AGC TGC AAC AGG CGG
2281
TTA CGG TGC AAA TGA CTA ATT TGC ACT GTT TCT AAA TCC ATT CCT CGT ATG TAT TCG AGT
2341
CGT TCT ATT ATT TTT AGA AAA GTT TCG GGT GAA GGA TGA CCC GGT GGC TCT TTT AAC CAA

2401
CCC AAT ATC GTT TTA TTG GAT TCG GAT GGA TGC TGC GAG GTA ATA ATC CCT TCA AGC TTT
2461
TCT TTT TGC TCA TTT GTT AGA GAT TTA CTA ACC GTA TTA AAT AGC TTC TTT TCA GCC ATT
2521
GCC CTT GCT TCC CAC ACC ATT CTT TCA AGT GTA GTG ATA GCA GGC AGT ATA ATT TTG TTT
2581
TTT CTT AGA AAA TCT ATG CAT TCA TGC AGT AGA TGA ATG GCA TCA CCA TTT TCC AAA GCT
2641
AAT TGA TGA AGG TAC TTA AAT GTC ATT CGA TAT TCA CTC AGG GTA AAA GTT ACA AAG TCG
2701
TAT TCA CTT CGA ATT TCT TTC AAA TGA TCC CAA AGT GTA TTT TCC CTT TGA GGA TAA TGA
2761
TCA AGC GAG GAT GGA CTA ACA CCA ATC TGT TTC GAT ATA TAT TGT ATG ACC GAA TCT GGG
2821
ATG CTT TTG ATA TGA GTG TAT GGC CAA CCG GGA TAC CGA AGA ACA GCT AAT TGA ACA GCA
2881
AAT CCT AAA CGG TTT TCT TCC CTC CTT CGC TTA TTA ACT ATT TCT AAA TCC CGT TTG GAA
2941
AAA GTG AAG TAG GTC CCC AGT ATC CAT TCA TCT TCA GGG ATT TGC ATA AAA GCC TGT CTC
3001
TGT TCC GGT GTA AGC AAT TCT CTA CCT CTC GCA ATT TTC ATT CAG TAT CAT TCC ATT TCT
3061
GTA TTT TCA ATT TAT TAG TTC AAT TAT ATA TCA ATA GAG TGT ACT CTA TTG ATA CAA ATG
3121
TAG TAG ACT GAT AAA ATC ATA GTT AAG AGC GTC TCA TAA GAC TTG TCT CAA AAA TGA GGT

3181 **résolvase**

LEU ARG LYS ILE GLY TYR ILE ARG VAL SER SER THR ASN GLN ASN PRO SER ARG
GAT ATT TTG CGG AAA ATC GGT TAT ATT CGT GTC AGT TCG ACT AAC CAG AAT CCT TCA AGA

3241

GLN PHE GLN GLN LEU ASN GLU ILE GLY MET ASP ILE ILE TYR GLU GLU LYS VAL SER GLY
CAA TTT CAG CAG TTG AAC GAG ATC GGA ATG GAT ATT ATA TAT GAA GAG AAA GTT TCA GGA

3301

ALA THR LYS ASP ARG GLU GLN LEU GLN LYS VAL LEU ASP ASP LEU GLN GLU ASP ASP ILE
GCA ACA AAG GAT CGC GAG CAA CTT CAA AAA GTG TTA GAC GAT TTA CAG GAA GAT GAC ATC

3361

ILE TYR VAL THR ASP LEU THR ARG ILE THR ARG SER THR GLN ASP LEU PHE GLU LEU ILE
ATT TAT GTT ACA GAC TTA ACT CGA ATC ACT CGT AGT ACA CAA GAT CTA TTT GAA TTA ATC

3421

ASP ASN ILE ARG ASP LYS LYS ALA SER LEU LYS SER LEU LYS ASP THR TRP LEU ASP LEU
GAT AAC ATA CGA GAT AAA AAG GCA AGT TTA AAA TCA CTA AAA GAT ACA TGG CTT GAT TTA

3481

SER GLU ASP ASN PRO TYR SER GLN PHE LEU ILE THR VAL MET ALA GLY VAL ASN GLN LEU
TCA GAA GAT AAT CCA TAC AGC CAA TTC TTA ATT ACT GTA ATG GCT GGT GTT AAC CAA TTA

3541

GLU ARG ASP LEU ILE ARG MET ARG GLN ARG GLU GLY ILE GLU LEU ALA LYS LYS GLU GLY
GAG CGA GAT CTT ATT CGG ATG AGA CAA CGT GAA GGG ATT GAA TTG GCT AAG AAA GAA GGA

3601

LYS PHE LYS GLY ARG LEU LYS LYS TYR HIS LYS ASN HIS ALA GLY MET ASN TYR ALA VAL
AAG TTT AAA GGT CGA TTA AAG AAG TAT CAT AAA AAT CAC GCA GGA ATG AAT TAT GCG GTA

3661

LYS LEU TYR LYS GLU GLY ASN MET THR VAL ASN GLN ILE CYS GLU ILE THR ASN VAL SER
AAG CTA TAT AAA GAA GGA AAT ATG ACT GTA AAT CAA ATT TGT GAA ATT ACT AAT GTA TCT

3721

ARG ALA SER LEU TYR ARG LYS LEU SER GLU VAL ASN ASN
AGG GCT TCA TTA TAC AGG AAA TTA TCA GAA GTG AAT AAT TAG CCA TTC TGT ATT CCG CTA

```
3781
ATG GGC AAT ATT TTT AAA GAA GAA AAG GAA ACT ATA AAA TAT TAA CAG CCT CCT AGC GAT
3841
GCC GAA AAG CCC TTT GAT AAA AAA AGA ATC ATC ATC TTA AGA AAT TCT TAG TCA TTT ATT
3901
ATG TAA ATG CTT ATA AAT TCG GCC CTA TAA TCT GAT AAA TTA TTA AGG GCA AAC TTA TGT

3961          VanR  MET SER ASP LYS ILE LEU ILE VAL ASP ASP GLU HIS GLU ILE ALA
GAA AGG GTG ATA ACT ATG AGC GAT AAA ATA CTT ATT GTG GAT GAT GAA CAT GAA ATT GCC
4021
ASP LEU VAL GLU LEU TYR LEU LYS ASN GLU ASN TYR THR VAL PHE LYS TYR TYR THR ALA
GAT TTG GTT GAA TTA TAC TTA AAA AAC GAG AAT TAT ACG GTT TTC AAA TAC TAT ACC GCC
4081
LYS GLU ALA LEU GLU CYS ILE ASP LYS SER GLU ILE ASP LEU ALA ILE LEU ASP ILE MET
AAA GAA GCA TTG GAA TGT ATA GAC AAG TCT GAG ATT GAC CTT GCC ATA TTG GAC ATC ATG
4141
LEU PRO GLY THR SER GLY LEU THR ILE CYS GLN LYS ILE ARG ASP LYS HIS THR TYR PRO
CTT CCC GGC ACA AGC GGC CTT ACT ATC TGT CAA AAA ATA AGG GAC AAG CAC ACC TAT CCG
4201
ILE ILE MET LEU THR GLY LYS ASP THR GLU VAL ASP LYS ILE THR GLY LEU THR ILE GLY
ATT ATC ATG CTG ACC GGG AAA GAT ACA GAG GTA GAT AAA ATT ACA GGG TTA ACA ATC GGC
4261
ALA ASP ASP TYR ILE THR LYS PRO PHE ARG PRO LEU GLU LEU ILE ALA ARG VAL LYS ALA
GCG GAT GAT TAT ATA ACG AAG CCC TTT CGC CCA CTG GAG TTA ATT GCT CGG GTA AAG GCC
4321
GLN LEU ARG ARG TYR LYS LYS PHE SER GLY VAL LYS GLU GLN ASN GLU ASN VAL ILE VAL
CAG TTG CGC CGA TAC AAA AAA TTC AGT GGA GTA AAG GAG CAG AAC GAA AAT GTT ATC GTC
```

```
4381
HIS SER GLY LEU VAL ILE ASN VAL ASN THR HIS GLU CYS TYR LEU ASN GLU LYS GLN LEU
CAC TCC GGC CTT GTC ATT AAT GTT AAC ACC CAT GAG TGT TAT CTG AAC GAG AAG CAG TTA
4441
SER LEU THR PRO THR GLU PHE SER ILE LEU ARG ILE LEU CYS GLU ASN LYS GLY ASN VAL
TCC CTT ACT CCC ACC GAG TTT TCA ATA CTG CGA ATC CTC TGT GAA AAC AAG GGG AAT GTG
4501
VAL SER SER GLU LEU LEU PHE HIS GLU ILE TRP GLY ASP GLU TYR PHE SER LYS SER ASN
GTT AGC TCC GAG CTG CTA TTT CAT GAG ATA TGG GGC GAC GAA TAT TTC AGC AAG AGC AAC
4561
ASN THR ILE THR VAL HIS ILE ARG HIS LEU ARG GLU LYS MET ASN ASP THR ILE ASP ASN
AAC ACC ATC ACC GTG CAT ATC CGG CAT TTG CGC GAA AAA ATG AAC GAC ACC ATT GAT AAT
4621
PRO LYS TYR ILE LYS THR VAL TRP GLY VALGLYTYRLYSILEGLULYS
CCG AAA TAT ATA AAA ACG GTA TGG GGG GTTGGTTATAAAATTGAAAAAT AAA AAA AAC GAC
                              VanS    LEUVALILELYSLEULYSASN LYS LYS ASN ASP
4682
TYR SER LYS LEU GLU ARG LYS LEU TYR MET TYR ILE VAL ALA ILE VAL VAL VAL ALA ILE
TAT TCC AAA CTA GAA CGA AAA CTT TAC ATG TAT ATC GTT GCA ATT GTT GTG GTA GCA ATT
4742
VAL PHE VAL LEU TYR ILE ARG SER MET ILE ARG GLY LYS LEU GLY ASP TRP ILE LEU SER
GTA TTC GTG TTG TAT ATT CGT TCA ATG ATC CGA GGG AAA CTT GGG GAT TGG ATC TTA AGT
4802
ILE LEU GLU ASN LYS TYR ASP LEU ASN HIS LEU ASP ALA MET LYS LEU TYR GLN TYR SER
ATT TTG GAA AAC AAA TAT GAC TTA AAT CAC CTG GAC GCG ATG AAA TTA TAT CAA TAT TCC
4862
ILE ARG ASN ASN ILE ASP ILE PHE ILE TYR VAL ALA ILE VAL ILE SER ILE LEU ILE LEU
ATA CGG AAC AAT ATA GAT ATC TTT ATT TAT GTG GCG ATT GTC ATT AGT ATT CTT ATT CTA
4922
CYS ARG VAL MET LEU SER LYS PHE ALA LYS TYR PHE ASP GLU ILE ASN THR GLY ILE ASP
TGT CGC GTC ATG CTT TCA AAA TTC GCA AAA TAC TTT GAC GAG ATA AAT ACC GGC ATT GAT
```

4982
VAL LEU ILE GLN ASN GLU ASP LYS GLN ILE GLU LEU SER ALA GLU MET ASP VAL MET GLU
GTA CTT ATT CAG AAC GAA GAT AAA CAA ATT GAG CTT TCT GCG GAA ATG GAT GTT ATG GAA
5042
GLN LYS LEU ASN THR LEU LYS ARG THR LEU GLU LYS ARG GLU GLN ASP ALA LYS LEU ALA
CAA AAG CTC AAC ACA TTA AAA CGG ACT CTG GAA AAG CGA GAG CAG GAT GCA AAG CTG GCC
5102
GLU GLN ARG LYS ASN ASP VAL VAL MET TYR LEU ALA HIS ASP ILE LYS THR PRO LEU THR
GAA CAA AGA AAA AAT GAC GTT GTT ATG TAC TTG GCG CAC GAT ATT AAA ACG CCC CTT ACA
5162
SER ILE ILE GLY TYR LEU SER LEU LEU ASP GLU ALA PRO ASP MET PRO VAL ASP GLN LYS
TCC ATT ATC GGT TAT TTG AGC CTG CTT GAC GAG GCT CCA GAC ATG CCG GTA GAT CAA AAG
5222
ALA LYS TYR VAL HIS ILE THR LEU ASP LYS ALA TYR ARG LEU GLU GLN LEU ILE ASP GLU
GCA AAG TAT GTG CAT ATC ACG TTG GAC AAA GCG TAT CGA CTC GAA CAG CTA ATC GAC GAG
5282
PHE PHE GLU ILE THR ARG TYR ASN LEU GLN THR ILE THR LEU THR LYS THR HIS ILE ASP
TTT TTT GAG ATT ACA CGG TAT AAC CTA CAA ACG ATA ACG CTA ACA AAA ACG CAC ATA GAC
5342
LEU TYR TYR MET LEU VAL GLN MET THR ASP GLU PHE TYR PRO GLN LEU SER ALA HIS GLY
CTA TAC TAT ATG CTG GTG CAG ATG ACC GAT GAA TTT TAT CCT CAG CTT TCC GCA CAT GGA
5402
LYS GLN ALA VAL ILE HIS ALA PRO GLU ASP LEU THR VAL SER GLY ASP PRO ASP LYS LEU
AAA CAG GCG GTT ATT CAC GCC CCC GAG GAT CTG ACC GTG TCC GGC GAC CCT GAT AAA CTC
5462
ALA ARG VAL PHE ASN ASN ILE LEU LYS ASN ALA ALA ALA TYR SER GLU ASP ASN SER ILE
GCG AGA GTC TTT AAC AAC ATT TTG AAA AAC GCC GCT GCA TAC AGT GAG GAT AAC AGC ATC

5522
ILE ASP ILE THR ALA GLY LEU SER GLY ASP VAL VAL SER ILE GLU PHE LYS ASN THR GLY
ATT GAC ATT ACC GCG GGC CTC TCC GGG GAT GTG GTG TCA ATC GAA TTC AAG AAC ACT GGA
5582
SER ILE PRO LYS ASP LYS LEU ALA ALA ILE PHE GLU LYS PHE TYR ARG LEU ASP ASN ALA
AGC ATC CCA AAA GAT AAG CTA GCT GCC ATA TTT GAA AAG TTC TAT AGG CTG GAC AAT GCT
5642
ARG SER SER ASP THR GLY GLY ALA GLY LEU GLY LEU ALA ILE ALA LYS GLU ILE ILE VAL
CGT TCT TCC GAT ACG GGT GGC GCG GGA CTT GGA TTG GCG ATT GCA AAA GAA ATT ATT GTT
5702
GLN HIS GLY GLY GLN ILE TYR ALA GLU SER ASN ASP ASN TYR THR THR PHE ARG VAL GLU
CAG CAT GGA GGG CAG ATT TAC GCG GAA AGC AAT GAT AAC TAT ACG ACG TTT AGG GTA GAG
5762
LEU PRO ALA MET PRO ASP LEU VAL ASP LYS ARG ARG SER
CTT CCA GCG ATG CCA GAC TTG GTT GAT AAA AGG AGG TCC TAA GA GAT GTA TAT AAT TTT
5821
TTA GGA AAA TCT CAA GGT TAT CTT TAC TTT TTC TTA GGA AAT TAA CAA TTT AAT ATT AAG
5881
AAA CGG CTC GTT CTT ACA CGG TAG ACT TAA TAC CGT AAG AAC GAG CCG TTT TCG TTC TTC
5941
AGA GAA AGA TTT GAC AAG ATT ACC ATT GGC ATC CCC GTT TTA TTT GGT GCC TTT CAC AGA
6001

VanH MET ASN ASN ILE GLY ILE THR VAL TYR GLY CYS GLU GLN ASP GLU
AAGGGTTGG TCT TAA TT ATG AAT AAC ATC GGC ATT ACT GTT TAT GGA TGT GAG CAG GAT GAG
6063
ALA ASP ALA PHE HIS ALA LEU SER PRO ARG PHE GLY VAL MET ALA THR ILE ILE ASN ALA
GCA GAT GCA TTC CAT GCT CTT TCG CCT CGC TTT GGC GTT ATG GCA ACG ATA ATT AAC GCC

6123
ASN VAL SER GLU SER ASN ALA LYS SER ALA PRO PHE ASN GLN CYS ILE SER VAL GLY HIS
AAC GTG TCG GAA TCC AAC GCC AAA TCC GCG CCT TTC AAT CAA TGT ATC AGT GTG GGA CAT
6183
LYS SER GLU ILE SER ALA SER ILE LEU LEU ALA LEU LYS ARG ALA GLY VAL LYS TYR ILE
AAA TCA GAG ATT TCC GCC TCT ATT CTT CTT GCG CTG AAG AGA GCC GGT GTG AAA TAT ATT
6243
SER THR ARG SER ILE GLY CYS ASN HIS ILE ASP THR THR ALA ALA LYS ARG MET GLY ILE
TCT ACC CGA AGC ATC GGC TGC AAT CAT ATA GAT ACA ACT GCT GCT AAG AGA ATG GGC ATC
6303
THR VAL ASP ASN VAL ALA TYR SER PRO ASP SER VAL ALA ASP TYR THR MET MET LEU ILE
ACT GTC GAC AAT GTG GCG TAC TCG CCG GAT AGC GTT GCC GAT TAT ACT ATG ATG CTA ATT
6363
LEU MET ALA VAL ARG ASN VAL LYS SER ILE VAL ARG SER VAL GLU LYS HIS ASP PHE ARG
CTT ATG GCA GTA CGC AAC GTA AAA TCG ATT GTG CGC TCT GTG GAA AAA CAT GAT TTC AGG
6423
LEU ASP SER ASP ARG GLY LYS VAL LEU SER ASP MET THR VAL GLY VAL VAL GLY THR GLY
TTG GAC AGC GAC CGT GGC AAG GTA CTC AGC GAC ATG ACA GTT GGT GTG GTG GGA ACG GGC
6483
GLN ILE GLY LYS ALA VAL ILE GLU ARG LEU ARG GLY PHE GLY CYS LYS VAL LEU ALA TYR
CAG ATA GGC AAA GCG GTT ATT GAG CGG CTG CGA GGA TTT GGA TGT AAA GTG TTG GCT TAT
6543
SER ARG SER ARG SER ILE GLU VAL ASN TYR VAL PRO PHE ASP GLU LEU LEU GLN ASN SER
AGT CGC AGC CGA AGT ATA GAG GTA AAC TAT GTA CCG TTT GAT GAG TTG CTG CAA AAT AGC
6603
ASP ILE VAL THR LEU HIS VAL PRO LEU ASN THR ASP THR HIS TYR ILE ILE SER HIS GLU
GAT ATC GTT ACG CTT CAT GTG CCG CTC AAT ACG GAT ACG CAC TAT ATT ATC AGC CAC GAA
6663
GLN ILE GLN ARG MET LYS GLN GLY ALA PHE LEU ILE ASN THR GLY ARG GLY PRO LEU VAL
CAA ATA CAG AGA ATG AAG CAA GGA GCA TTT CTT ATC AAT ACT GGG CGC GGT CCA CTT GTA

6723
ASP THR TYR GLU LEU VAL LYS ALA LEU GLU ASN GLY LYS LEU GLY GLY ALA ALA LEU ASP
GAT ACC TAT GAG TTG GTT AAA GCA TTA GAA AAC GGG AAA CTG GGC GGT GCC GCA TTG GAT
6783
VAL LEU GLU GLY GLU GLU GLU PHE PHE TYR SER ASP CYS THR GLN LYS PRO ILE ASP ASN
GTA TTG GAA GGA GAG GAA GAG TTT TTC TAC TCT GAT TGC ACC CAA AAA CCA ATT GAT AAT
6843
GLN PHE LEU LEU LYS LEU GLN ARG MET PRO ASN VAL ILE ILE THR PRO HIS THR ALA TYR
CAA TTT TTA CTT AAA CTT CAA AGA ATG CCT AAC GTG ATA ATC ACA CCG CAT ACG GCC TAT
6903
TYR THR GLU GLN ALA LEU ARG ASP THR VAL GLU LYS THR ILE LYS ASN CYS LEU ASP PHE
TAT ACC GAG CAA GCG TTG CGT GAT ACC GTT GAA AAA ACC ATT AAA AAC TGT TTG GAT TTT
6963
VanA METASN ARG ILE LYS VAL ALA ILE LEU PHE GLY GLY CYS SER
GAA AGG AGA CAG GAG CATGAAT AGA ATA AAA GTT GCA ATA CTG TTT GGG GGT TGC TCA
GLU ARG ARG GLN GLU HISGLU
7021
GLU GLU HIS ASP VAL SER VAL LYS SER ALA ILE GLU ILE ALA ALA ASN ILE ASN LYS GLU
GAG GAG CAT GAC GTA TCG GTA AAA TCT GCA ATA GAG ATA GCC GCT AAC ATT AAT AAA GAA
7081
LYS TYR GLU PRO LEU TYR ILE GLY ILE THR LYS SER GLY VAL TRP LYS MET CYS GLU LYS
AAA TAC GAG CCG TTA TAC ATT GGA ATT ACG AAA TCT GGT GTA TGG AAA ATG TGC GAA AAA
7141
PRO CYS ALA GLU TRP GLU ASN ASP ASN CYS TYR SER ALA VAL LEU SER PRO ASP LYS LYS
CCT TGC GCG GAA TGG GAA AAC GAC AAT TGC TAT TCA GCT GTA CTC TCG CCG GAT AAA AAA
7201
MET HIS GLY LEU LEU VAL LYS LYS ASN HIS GLU TYR GLU ILE ASN HIS VAL ASP VAL ALA
ATG CAC GGA TTA CTT GTT AAA AAG AAC CAT GAA TAT GAA ATC AAC CAT GTT GAT GTA GCA

7261
PHE SER ALA LEU HIS GLY LYS SER GLY GLU ASP GLY SER ILE GLN GLY LEU PHE GLU LEU
TTT TCA GCT TTG CAT GGC AAG TCA GGT GAA GAT GGA TCC ATA CAA GGT CTG TTT GAA TTG
7321
SER GLY ILE PRO PHE VAL GLY CYS ASP ILE GLN SER SER ALA ILE CYS MET ASP LYS SER
TCC GGT ATC CCT TTT GTA GGC TGC GAT ATT CAA AGC TCA GCA ATT TGT ATG GAC AAA TCG
7381
LEU THR TYR ILE VAL ALA LYS ASN ALA GLY ILE ALA THR PRO ALA PHE TRP VAL ILE ASN
TTG ACA TAC ATC GTT GCG AAA AAT GCT GGG ATA GCT ACT CCC GCC TTT TGG GTT ATT AAT
7441
LYS ASP ASP ARG PRO VAL ALA ALA THR PHE THR TYR PRO VAL PHE VAL LYS PRO ALA ARG
AAA GAT GAT AGG CCG GTG GCA GCT ACG TTT ACC TAT CCT GTT TTT GTT AAG CCG GCG CGT
7501
SER GLY SER SER PHE GLY VAL LYS LYS VAL ASN SER ALA ASP GLU LEU ASP TYR ALA ILE
TCA GGC TCA TCC TTC GGT GTG AAA AAA GTC AAT AGC GCG GAC GAA TTG GAC TAC GCA ATT
7561
GLU SER ALA ARG GLN TYR ASP SER LYS ILE LEU ILE GLU GLN ALA VAL SER GLY CYS GLU
GAA TCG GCA AGA CAA TAT GAC AGC AAA ATC TTA ATT GAG CAG GCT GTT TCG GGC TGT GAG
7621
VAL GLY CYS ALA VAL LEU GLY ASN SER ALA ALA LEU VAL VAL GLY GLU VAL ASP GLN ILE
GTC GGT TGT GCG GTA TTG GGA AAC AGT GCC GCG TTA GTT GTT GGC GAG GTG GAC CAA ATC
7681
ARG LEU GLN TYR GLY ILE PHE ARG ILE HIS GLN GLU VAL GLU PRO GLU LYS GLY SER GLU
AGG CTG CAG TAC GGA ATC TTT CGT ATT CAT CAG GAA GTC GAG CCG GAA AAA GGC TCT GAA
7741
ASN ALA VAL ILE THR VAL PRO ALA ASP LEU SER ALA GLU GLU ARG GLY ARG ILE GLN GLU
AAC GCA GTT ATA ACC GTT CCC GCA GAC CTT TCA GCA GAG GAG CGA GGA CGG ATA CAG GAA
7801
THR ALA LYS LYS ILE TYR LYS ALA LEU GLY CYS ARG GLY LEU ALA ARG VAL ASP MET PHE
ACG GCA AAA AAA ATA TAT AAA GCG CTC GGC TGT AGA GGT CTA GCC CGT GTG GAT ATG TTT

```
7861
LEU GLN ASP ASN GLY ARG ILE VAL LEU ASN GLU VAL ASN THR LEU PRO GLY PHE THR SER
TTA CAA GAT AAC GGC CGC ATT GTA CTG AAC GAA GTC AAT ACT CTG CCC GGT TTC ACG TCA
7921
TYR SER ARG TYR PRO ARG MET MET ALA ALA ALA GLY ILE ALA LEU PRO GLU LEU ILE ASP
TAC AGT CGT TAT CCC CGT ATG ATG GCC GCT GCA GGT ATT GCA CTT CCC GAA CTG ATT GAC
7981
ARG LEU ILE VAL LEU ALA LEU LYS GLY
CGC TTG ATC GTA TTA GCG TTA AAG GGG TGATAAGC ATG GAA ATA GGA TTT ACT TTT TTA GAT
                                    VanX    MET GLU ILE GLY PHE THR PHE LEU ASP

8043
GLU ILE VAL HIS GLY VAL ARG TRP ASP ALA LYS TYR ALA THR TRP ASP ASN PHE THR GLY
GAA ATA GTA CAC GGT GTT CGT TGG GAC GCT AAA TAT GCC ACT TGG GAT AAT TTC ACC GGA
8103
LYS PRO VAL ASP GLY TYR GLU VAL ASN ARG ILE VAL GLY THR TYR GLU LEU ALA GLU SER
AAA CCG GTT GAC GGT TAT GAA GTA AAT CGC ATT GTA GGG ACA TAC GAG TTG GCT GAA TCG
8163
LEU LEU LYS ALA LYS GLU LEU ALA ALA THR GLN GLY TYR GLY LEU LEU LEU TRP ASP GLY
CTT TTG AAG GCA AAA GAA CTG GCT GCT ACC CAA GGG TAC GGA TTG CTT CTA TGG GAC GGT
8223
TYR ARG PRO LYS ARG ALA VAL ASN CYS PHE MET GLN TRP ALA ALA GLN PRO GLU ASN ASN
TAC CGT CCT AAG CGT GCT GTA AAC TGT TTT ATG CAA TGG GCT GCA CAG CCG GAA AAT AAC
8283
LEU THR LYS GLU SER TYR TYR PRO ASN ILE ASP ARG THR GLU MET ILE SER LYS GLY TYR
CTG ACA AAG GAA AGT TAT TAT CCC AAT ATT GAC CGA ACT GAG ATG ATT TCA AAA GGA TAC
8343
VAL ALA SER LYS SER SER HIS SER ARG GLY SER ALA ILE ASP LEU THR LEU TYR ARG LEU
GTG GCT TCA AAA TCA AGC CAT AGC CGC GGC AGT GCC ATT GAT CTT ACG CTT TAT CGA TTA
```

8403
ASP THR GLY GLU LEU VAL PRO MET GLY SER ARG PHE ASP PHE MET ASP GLU ARG SER HIS
GAC ACG GGT GAG CTT GTA CCA ATG GGG AGC CGA TTT GAT TTT ATG GAT GAA CGC TCT CAT
8463
HIS ALA ALA ASN GLY ILE SER CYS ASN GLU ALA GLN ASN ARG ARG ARG LEU ARG SER ILE
CAT GCG GCA AAT GGA ATA TCA TGC AAT GAA GCG CAA AAT CGC AGA CGT TTG CGC TCC ATC
8523
MET GLU ASN SER GLY PHE GLU ALA TYR SER LEU GLU TRP TRP HIS TYR VAL LEU ARG ASP
ATG GAA AAC AGT GGG TTT GAA GCA TAT AGC CTC GAA TGG TGG CAC TAT GTA TTA AGA GAC
8583
GLU PRO TYR PRO ASN SER TYR PHE ASP PHE PRO VAL LYS
GAA CCA TAC CCC AAT AGC TAT TTT GAT TTC CCC GTT AAA TAAA CTT TTA ACC GTT GCA
8641
CGG ACA AAC TAT ATA AGC TAA CTC TTT CGG CAG GAA ACC CGA CGT ATG TAA CTG GTT CTT
8701
AGG GAA TTT ATA TAT AGT AGA TAG TAT TGA AGA TGT AAG GCA GAG CGA TAT TGC GGT CAT
8761
TAT CTG CGT GCG CTG CGG CAA GAT AGC CTG ATA ATA AGA CTG ATC GCA TAG AGG GGT GGT
8821
ATT TCA CAC CGC CCA TTG TCA ACA GGC AGT TCA GCC TCG TTA AAT TCA GCA TGG GTA TCA
8881
CTT ATG AAA ATT CAT CTA CAT TGG TGA TAA TAG TAA ATC CAG TAG GGC GAA ATA ATT GAC
8941
TGT AAT TTA CGG GGC AAA ACG GCA CAA TCT CAA ACG AGA TTG TGC CGT TTA AGG GGA AGA
9001
VanY MET LYS LYS
TTC TAG AAA TAT TTC ATA CTT CCA ACT ATA TAG TTA AGG AGG AGA CTG AAA ATG AAG AAG
9061
LEU PHE PHE LEU LEU LEU LEU LEU PHE LEU ILE TYR LEU GLY TYR ASP TYR VAL ASN GLU
TTG TTT TTT TTA TTG TTA TTG TTA TTC TTA ATA TAC TTA GGT TAT GAC TAC GTT AAT GAA

```
9121
ALA LEU PHE SER GLN GLU LYS VAL GLU PHE GLN ASN TYR ASP GLN ASN PRO LYS GLU HIS
GCA CTG TTT TCT CAG GAA AAA GTC GAA TTT CAA AAT TAT GAT CAA AAT CCC AAA GAA CAT
9181
LEU GLU ASN SER GLY THR SER GLU ASN THR GLN GLU LYS THR ILE THR GLU GLU GLN VAL
TTA GAA AAT AGT GGG ACT TCT GAA AAT ACC CAA GAG AAA ACA ATT ACA GAA GAA CAG GTT
9241
TYR GLN GLY ASN LEU LEU LEU ILE ASN SER LYS TYR PRO VAL ARG GLN GLU SER VAL LYS
TAT CAA GGA AAT CTG CTA TTA ATC AAT AGT AAA TAT CCT GTT CGC CAA GAA AGT GTG AAG
9301
SER ASP ILE VAL ASN LEU SER LYS HIS ASP GLU LEU ILE ASN GLY TYR GLY LEU LEU ASP
TCA GAT ATC GTG AAT TTA TCT AAA CAT GAC GAA TTA ATA AAT GGA TAC GGG TTG CTT GAT
9361
SER ASN ILE TYR MET SER LYS GLU ILE ALA GLN LYS PHE SER GLU MET VAL ASN ASP ALA
AGT AAT ATT TAT ATG TCA AAA GAA ATA GCA CAA AAA TTT TCA GAG ATG GTC AAT GAT GCT
9421
VAL LYS GLY GLY VAL SER HIS PHE ILE ILE ASN SER GLY TYR ARG ASP PHE ASP GLU GLN
GTA AAG GGT GGC GTT AGT CAT TTT ATT ATT AAT AGT GGC TAT CGA GAC TTT GAT GAG CAA
9481
SER VAL LEU TYR GLN GLU MET GLY ALA GLU TYR ALA LEU PRO ALA GLY TYR SER GLU HIS
AGT GTG CTT TAC CAA GAA ATG GGG GCT GAG TAT GCC TTA CCA GCA GGT TAT AGT GAG CAT
9541
ASN SER GLY LEU SER LEU ASP VAL GLY SER SER LEU THR LYS MET GLU ARG ALA PRO GLU
AAT TCA GGT TTA TCA CTA GAT GTA GGA TCA AGC TTG ACG AAA ATG GAA CGA GCC CCT GAA
9601
GLY LYS TRP ILE GLU GLU ASN ALA TRP LYS TYR GLY PHE ILE LEU ARG TYR PRO GLU ASP
GGA AAG TGG ATA GAA GAA AAT GCT TGG AAA TAC GGG TTC ATT TTA CGT TAT CCA GAG GAC
9661
LYS THR GLU LEU THR GLY ILE GLN TYR GLU PRO TRP HIS ILE ARG TYR VAL GLY LEU PRO
AAA ACA GAG TTA ACA GGA ATT CAA TAT GAA CCA TGG CAT ATT CGC TAT GTT GGT TTA CCA
```

```
9721
HIS SER ALA ILE MET LYS GLU LYS ASN PHE VAL LEU GLU GLU TYR MET ASP TYR LEU LYS
CAT AGT GCG ATT ATG AAA GAA AAG AAT TTC GTT CTC GAG GAA TAT ATG GAT TAC CTA AAA
9781
GLU GLU LYS THR ILE SER VAL SER VAL ASN GLY GLU LYS TYR GLU ILE PHE TYR TYR PRO
GAA GAA AAA ACC ATT TCT GTT AGT GTA AAT GGG GAA AAA TAT GAG ATC TTT TAT TAT CCT
9841
VAL THR LYS ASN THR THR ILE HIS VAL PRO THR ASN LEU ARG TYR GLU ILE SER GLY ASN
GTT ACT AAA AAT ACC ACC ATT CAT GTG CCG ACT AAT CTT CGT TAT GAG ATA TCA GGA AAC
9901
ASN ILE ASP GLY VAL ILE VAL THR VAL PHE PRO GLY SER THR HIS THR ASN SER ARG ARG
AAT ATA GAC GGT GTA ATT GTG ACA GTG TTT CCC GGA TCA ACA CAT ACT AAT TCA AGG AGG
9961
TAA GGA TGG CGG AAT GAA ACC AAC GAA ATT AAT GAA CAG CAT TAT TGT ACT AGC ACT TTT
10021
GGG GTA ACG TTA GCT TTT TAA TTT AAA ACC CAC GTT AAC TAG GAC ATT GCT ATA CTA ATG

10081                                  VanZ     LEU GLY LYS ILE LEU SER ARG GLY LEU
ATA CAA CTT AAA CAA AAG AATTAGAGG AAA TTA TA TTG GGA AAA ATA TTA TCT AGA GGA TTG
10143
LEU ALA LEU TYR LEU VAL THR LEU ILE TRP LEU VAL LEU PHE LYS LEU GLN TYR ASN ILE
CTA GCT TTA TAT TTA GTG ACA CTA ATC TGG TTA GTG TTA TTC AAA TTA CAA TAC AAT ATT
10203
LEU SER VAL PHE ASN TYR HIS GLN ARG SER LEU ASN LEU THR PRO PHE THR ALA THR GLY
TTA TCA GTA TTT AAT TAT CAT CAA AGA AGT CTT AAC TTG ACT CCA TTT ACT GCT ACT GGG
10263
ASN PHE ARG GLU MET ILE ASP ASN VAL ILE ILE PHE ILE PRO PHE GLY LEU LEU LEU ASN
AAT TTC AGA GAG ATG ATA GAT AAT GTT ATA ATC TTT ATT CCA TTT GGC TTG CTT TTG AAT
```

10323
VAL ASN PHE LYS GLU ILE GLY PHE LEU PRO LYS PHE ALA PHE VAL LEU VAL LEU SER LEU
GTC AAT TTT AAA GAA ATC GGA TTT TTA CCT AAG TTT GCT TTT GTA CTG GTT TTA AGT CTT
10383
THR PHE GLU ILE ILE GLN PHE ILE PHE ALA ILE GLY ALA THR ASP ILE THR ASP VAL ILE
ACT TTT GAA ATA ATT CAA TTT ATC TTC GCT ATT GGA GCG ACA GAC ATA ACA GAT GTA ATT
10443
THR ASN THR VAL GLY GLY PHE LEU GLY LEU LYS LEU TYR GLY LEU SER ASN LYS HIS MET
ACA AAT ACT GTT GGA GGC TTT CTT GGA CTG AAA TTA TAT GGT TTA AGC AAT AAG CAT ATG
10503
ASN GLN LYS LYS LEU ASP ARG VAL ILE ILE PHE VAL GLY ILE LEU LEU LEU VAL LEU LEU
AAT CAA AAA AAA TTA GAC AGA GTT ATT ATT TTT GTA GGT ATA CTT TTG CTC GTA TTA TTG
10563
LEU VAL TYR ARG THR HIS LEU ARG ILE ASN TYR VAL
CTC GTT TAC CGT ACC CAT TTA AGA ATA AAT TAC GTG TAAG ATG TCT AAA TCA AGC AAT
10621
CTG ATC TTT CAT ACA CAT AAA GAT ATT GAA TGA ATT GGA TTA GAT GGA AAA CGG GAT GTG
10681
GGG AAA CTC GCC CGT AGG TGT GAA GTG AGG GGA AAA CCG GTG ATA AAG TAA AAA GCT TAC
10741
CTA ACA CTA TAG TAA CAA AGA AAG CCC AAT TAT CAA TTT TAG TGC TGA GGA ATT GGT CTC
10801
TTT AAT AAA TTT CCT TAA CGT TGT AAA TCC GCA TTT TCC TGA CGG TAC CCC

Ib (-) Strand

(corresponds to the sequence of the strand complementary to the (+) strand from 1 to 3189.

```
1
CAA AAT ATC ACC TCA TTT TTG AGA CAA GTC TTA TGA GAC GCT CTT AAC TAT GAT TTT ATC
61
AGT CTA CTA CAT TTG TAT CAA TAG AGT ACA CTC TAT TGA TAT ATA ATT GAA CTA ATA AAT

121              Transposase          MET LYS ILE ALA ARG GLY ARG GLU LEU LEU THR
TGA AAA TAC AGA AAT GGA ATA TACTG AA ATG AAA ATT GCG AGA GGT AGA GAA TTG CTT ACA
182
PRO GLU GLN ARG GLN ALA PHE MET GLN ILE PRO GLU ASP GLU TRP ILE LEU GLY THR TYR
CCG GAA CAG AGA CAG GCT TTT ATG CAA ATC CCT GAA GAT GAA TGG ATA CTG GGG ACC TAC
242
PHE THR PHE SER LYS ARG ASP LEU GLU ILE VAL ASN LYS ARG ARG ARG GLU GLU ASN ARG
TTC ACT TTT TCC AAA CGG GAT TTA GAA ATA GTT AAT AAG CGA AGG AGG GAA GAA AAC CGT
302
LEU GLY PHE ALA VAL GLN LEU ALA VAL LEU ARG TYR PRO GLY TRP PRO TYR THR HIS ILE
TTA GGA TTT GCT GTT CAA TTA GCT GTT CTT CGG TAT CCC GGT TGG CCA TAC ACT CAT ATC
362
LYS SER ILE PRO ASP SER VAL ILE GLN TYR ILE SER LYS GLN ILE GLY VAL SER PRO SER
AAA AGC ATC CCA GAT TCG GTC ATA CAA TAT ATA TCG AAA CAG ATT GGT GTT AGT CCA TCC
422
SER LEU ASP HIS TYR PRO GLN ARG GLU ASN THR LEU TRP ASP HIS LEU LYS GLU ILE ARG
TCG CTT GAT CAT TAT CCT CAA AGG GAA AAT ACA CTT TGG GAT CAT TTG AAA GAA ATT CGA
```

```
482
SER GLU TYR ASP PHE VAL THR PHE THR LEU SER GLU TYR ARG MET THR PHE LYS TYR LEU
AGT GAA TAC GAC TTT GTA ACT TTT ACC CTG AGT GAA TAT CGA ATG ACA TTT AAG TAC CTT
542
HIS GLN LEU ALA LEU GLU ASN GLY ASP ALA ILE HIS LEU LEU HIS GLU CYS ILE ASP PHE
CAT CAA TTA GCT TTG GAA AAT GGT GAT GCC ATT CAT CTA CTG CAT GAA TGC ATA GAT TTT
602
LEU ARG LYS ASN LYS ILE ILE LEU PRO ALA ILE THR THR LEU GLU ARG MET VAL TRP GLU
CTA AGA AAA AAC AAA ATT ATA CTG CCT GCT ATC ACT ACA CTT GAA AGA ATG GTG TGG GAA
662
ALA ARG ALA MET ALA GLU LYS LYS LEU PHE ASN THR VAL SER LYS SER LEU THR ASN GLU
GCA AGG GCA ATG GCT GAA AAG AAG CTA TTT AAT ACG GTT AGT AAA TCT CTA ACA AAT GAG
722
GLN LYS GLU LYS LEU GLU GLY ILE ILE THR SER GLN HIS PRO SER GLU SER ASN LYS THR
CAA AAA GAA AAG CTT GAA GGG ATT ATT ACC TCG CAG CAT CCA TCC GAA TCC AAT AAA ACG
782
ILE LEU GLY TRP LEU LYS GLU PRO PRO GLY HIS PRO SER PRO GLU THR PHE LEU LYS ILE
ATA TTG GGT TGG TTA AAA GAG CCA CCG GGT CAT CCT TCA CCC GAA ACT TTT CTA AAA ATA
842
ILE GLU ARG LEU GLU TYR ILE ARG GLY MET ASP LEU GLU THR VAL GLN ILE SER HIS LEU
ATA GAA CGA CTC GAA TAC ATA CGA GGA ATG GAT TTA GAA ACA GTG CAA ATT AGT CAT TTG
902
HIS ARG ASN ARG LEU LEU GLN LEU SER ARG LEU GLY SER ARG TYR GLU PRO TYR ALA PHE
CAC CGT AAC CGC CTG TTG CAG CTG TCT CGC TTA GGC TCA AGA TAC GAG CCG TAT GCA TTC
962
ARG ASP PHE GLN GLU ASN LYS ARG TYR SER ILE LEU THR ILE TYR LEU LEU GLN LEU THR
CGT GAC TTT CAA GAA AAT AAA CGT TAT TCG ATA TTA ACC ATC TAT TTA TTA CAA CTT ACT
```

1022
GLN GLU LEU THR ASP LYS ALA PHE GLU ILE HIS ASP ARG GLN ILE LEU SER LEU LEU SER
CAG GAG CTA ACG GAT AAA GCG TTT GAA ATT CAT GAT AGG CAA ATA CTT AGT TTG TTA TCA
1082
LYS GLY ARG LYS ALA GLN GLU GLU ILE GLN LYS GLN ASN GLY LYS LYS LEU ASN GLU LYS
AAA GGT CGT AAG GCT CAA GAG GAA ATC CAG AAA CAA AAC GGT AAA AAG CTA AAT GAG AAA
1142
VAL ILE HIS PHE THR ASN ILE GLY GLN ALA LEU ILE LYS ALA ARG GLU GLU LYS LEU ASP
GTT ATA CAC TTT ACG AAC ATC GGA CAA GCA TTA ATT AAA GCA AGA GAG GAA AAA TTA GAC
1202
VAL PHE LYS VAL LEU GLU SER VAL ILE GLU TRP ASN THR PHE VAL SER SER VAL GLU GLU
GTT TTT AAG GTT TTA GAA TCG GTT ATT GAA TGG AAT ACC TTT GTC TCT TCA GTA GAA GAG
1262
ALA GLN GLU LEU ALA ARG PRO ALA ASP TYR ASP TYR LEU ASP LEU LEU GLN LYS ARG PHE
GCT CAG GAA CTT GCA CGT CCT GCC GAC TAT GAT TAT TTA GAC TTA CTG CAA AAA CGG TTT
1322
TYR SER LEU ARG LYS TYR THR PRO THR LEU LEU ARG VAL LEU GLU PHE HIS SER THR LYS
TAT TCA CTA AGA AAA TAT ACG CCA ACG CTA TTA AGA GTA TTG GAA TTT CAT TCT ACA AAG
1382
ALA ASN GLU PRO LEU LEU GLN ALA VAL GLU ILE ILE ARG GLY MET ASN GLU SER GLY LYS
GCA AAT GAG CCA CTT TTA CAA GCT GTT GAG ATT ATC CGA GGA ATG AAC GAA TCT GGA AAG
1442
ARG LYS VAL PRO ASP ASP SER PRO VAL ASP PHE ILE SER LYS ARG TRP LYS ARG HIS LEU
CGA AAA GTG CCT GAT GAC TCA CCT GTG GAT TTT ATT TCA AAA CGA TGG AAA AGA CAT TTA
1502
TYR GLU ASP ASP GLY THR THR ILE ASN ARG HIS TYR TYR GLU MET ALA VAL LEU THR GLU
TAC GAG GAT GAT GGT ACA ACA ATT AAT CGT CAT TAC TAT GAA ATG GCT GTT TTA ACA GAA
1562
LEU ARG GLU HIS VAL ARG ALA GLY ASP VAL SER ILE VAL GLY SER ARG GLN TYR ARG ASP
CTT CGG GAG CAT GTT CGG GCA GGA GAT GTT TCC ATT GTT GGC AGC AGA CAA TAT AGG GAT

1622
PHE GLU GLU TYR LEU PHE SER GLU ASP THR TRP ASN GLN SER LYS GLY ASN THR ARG LEU
TTT GAG GAA TAT TTG TTT TCG GAA GAT ACA TGG AAT CAA TCG AAG GGG AAT ACG AGA TTA
1682
SER VAL SER LEU SER PHE GLU ASP TYR ILE THR GLU ARG THR SER SER PHE ASN GLU ARG
TCA GTT AGT TTA TCA TTC GAA GAT TAT ATA ACG GAG AGA ACC AGC AGC TTT AAT GAA AGG
1742
LEU LYS TRP LEU ALA ALA ASN SER ASN LYS LEU ASP GLY VAL SER LEU GLU LYS GLY LYS
TTA AAG TGG TTA GCT GCC AAT TCC AAT AAG TTA GAT GGG GTT TCT CTT GAA AAA GGA AAG
1802
LEU SER LEU ALA ARG LEU GLU LYS ASP VAL PRO GLU GLU ALA LYS LYS PHE SER ALA SER
CTA TCA CTT GCA CGC TTA GAA AAA GAT GTT CCA GAA GAA GCA AAA AAA TTT AGT GCA AGC
1862
LEU TYR GLN MET LEU PRO ARG ILE LYS LEU THR ASP LEU LEU MET ASP VAL ALA HIS ILE
CTT TAT CAG ATG CTA CCA AGA ATA AAA TTA ACT GAT TTA CTC ATG GAT GTG GCC CAT ATA
1922
THR GLY PHE HIS GLU GLN PHE THR HIS ALA SER ASN ASN ARG LYS PRO ASP LYS GLU GLU
ACA GGA TTT CAT GAG CAA TTC ACT CAT GCT TCC AAT AAT CGA AAA CCA GAT AAG GAA GAA
1982
THR ILE ILE ILE MET ALA ALA LEU LEU GLY MET GLY MET ASN ILE GLY LEU SER LYS MET
ACA ATC ATT ATC ATG GCT GCC CTT TTA GGA ATG GGA ATG AAT ATT GGC TTG AGC AAG ATG
2042
ALA GLU ALA THR PRO GLY LEU THR TYR LYS GLN LEU ALA ASN VAL SER GLN TRP ARG MET
GCC GAA GCC ACA CCC GGA CTT ACA TAT AAG CAA CTA GCC AAT GTA TCT CAA TGG CGC ATG
2102
TYR GLU ASP ALA MET ASN LYS ALA GLN ALA ILE LEU VAL ASN PHE HIS HIS LYS LEU GLN
TAT GAA GAT GCC ATG AAT AAA GCC CAA GCC ATA TTA GTA AAC TTT CAT CAT AAA TTA CAA
2162
LEU PRO PHE TYR TRP GLY ASP GLY THR THR SER SER SER ASP GLY MET ARG MET GLN LEU
TTG CCT TTC TAT TGG GGC GAC GGT ACA ACA TCT TCG TCA GAT GGT ATG AGA ATG CAG CTA

2222
GLY VAL SER SER LEU HIS ALA ASP ALA ASN PRO HIS TYR GLY THR GLY LYS GLY ALA THR
GGT GTT TCA TCA CTA CAT GCA GAT GCA AAT CCA CAT TAT GGA ACT GGA AAA GGA GCC ACC
2282
ILE TYR ARG PHE THR SER ASP GLN PHE SER SER TYR TYR THR LYS ILE ILE HIS THR ASN
ATC TAC CGA TTT ACA AGT GAT CAA TTC TCT TCT TAC TAC ACA AAG ATT ATT CAT ACT AAT
2342
SER ARG ASP ALA ILE HIS VAL LEU ASP GLY LEU LEU HIS HIS GLU THR ASP LEU ASN ILE
TCA AGA GAT GCG ATT CAT GTT TTG GAT GGT TTG TTA CAT CAT GAG ACG GAT CTA AAC ATA
2402
GLU GLU HIS TYR THR ASP THR ALA GLY TYR THR ASP GLN ILE PHE GLY LEU THR HIS LEU
GAG GAA CAT TAT ACA GAC ACT GCC GGT TAC ACT GAC CAA ATA TTC GGA CTG ACT CAT TTA
2462
LEU GLY PHE LYS PHE ALA PRO ARG ILE ARG ASP LEU SER ASP SER LYS LEU PHE THR ILE
TTA GGA TTT AAA TTT GCC CCA AGA ATA AGG GAT TTA TCG GAC TCA AAA TTA TTT ACG ATA
2522
ASP LYS ALA SER GLU TYR PRO LYS LEU GLU ALA ILE LEU ARG GLY GLN ILE ASN THR LYS
GAT AAA GCA AGT GAG TAT CCA AAA CTA GAA GCC ATT TTA CGT GGA CAA ATA AAT ACA AAG
2582
VAL ILE LYS GLU ASN TYR GLU ASP VAL LEU ARG LEU ALA HIS SER ILE ARG GLU GLY THR
GTC ATT AAA GAA AAT TAT GAG GAT GTT TTG CGA TTA GCT CAT TCT ATA AGG GAG GGA ACA
2642
AGT TTC AGC ATC CCT TAT TAT GGG GAA GCT AGG TTC CTA TTC AAG ACA AAA CAG CTT AGC
VAL SER ALA SER LEU ILE MET GLY LYS LEU GLY SER TYR SER ARG GLN ASN SER LEU ALA
GTT TCA GCA TCC CTT ATT ATG GGG AAG CTA GGT TCC TAT TCA AGA CAA AAC AGC TTA GCT
2702
THR ALA LEU ARG GLU MET GLY ARG ILE GLU LYS THR ILE PHE ILE LEU ASN TYR ILE SER
ACA GCC TTA CGT GAG ATG GGC CGA ATA GAA AAA ACG ATC TTT ATT TTG AAT TAT ATA TCG

2762
ASP GLU SER LEU ARG ARG LYS ILE GLN ARG GLY LEU ASN LYS GLY GLU ALA MET ASN GLY
GAT GAA TCA TTA AGA AGA AAA ATA CAA AGA GGA TTG AAT AAA GGA GAA GCC ATG AAT GGA
2822
LEU ALA ARG ALA ILE PHE PHE GLY LYS GLN GLY GLU LEU ARG GLU ARG THR ILE GLN HIS
TTG GCA AGA GCT ATT TTC TTC GGA AAA CAA GGT GAG CTT AGA GAA CGC ACC ATA CAG CAT
2882
GLN LEU GLN ARG ALA SER ALA LEU ASN ILE ILE ILE ASN ALA ILE SER ILE TRP ASN THR
CAA TTG CAA AGA GCC AGT GCT TTA AAC ATA ATT ATC AAT GCT ATA AGT ATT TGG AAT ACT
2942
TCT CCA CCT AAC AAC AGC AGT TGA ATA TAA AAA ACG GAC AGG TAG CTT TAA TGA AGA TTT
LEU HIS LEU THR THR ALA VAL GLU TYR LYS LYS ARG THR GLY SER PHE ASN GLU ASP LEU
CTC CAC CTA ACA ACA GCA GTT GAA TAT AAA AAA CGG ACA GGT AGC TTT AAT GAA GAT TTG
3002
LEU HIS HIS MET SER PRO LEU GLY TRP GLU HIS ILE ASN LEU LEU GLY GLU TYR HIS PHE
TTA CAC CAT ATG TCG CCC TTA GGT TGG GAA CAT ATT AAT TTA CTA GGA GAA TAC CAT TTT
3062
ASN SER GLU LYS VAL VAL SER LEU ASN SER LEU ARG PRO LEU LYS LEU SER
AAC TCA GAG AAA GTA GTC TCA TTA AAT TCT TTA AGA CCA CTA AAA CTT TCT TAA CGT TG
3121
TTA AAA ACG AGG GAT TCG TCA GGA AAA TAG GCT TAG CGT TGT AAA TCC GCA TTT TCC TGA
3181
CGC TAC CCC

*Sac*I

```
GAGCTCTTCCTTCAACGCACTTCTGTACCAAGAGTTGTTGTC                                42

CATTTGATCACTAACAATAGCTTCCCCTGCTTTCTTCAAGCCCTTTGTCATAAAATCGTTAGATTTTCA    111

TCATAAAAATACGAGAAAGACAACAGGAAGACCGCAAATTTTCTTTTCTTTTCCTAGGTACACTGAATG    180

                          RBS                M  K  K  I  A  V  L  F  G  G
TAACCTTAAAAGAAAAAAGGAAAGGAAGAAAATGATGAAAAAAATTGCCGTTTTATTTGGAGGG         244

 N  S  E  E  Y  S  V  S  L  T  S  A  A  S  V  I  Q  A  I  D
AATTCTCCAGAATACTCAGTGTCACTAACCTCAGCAGCAAGTGTGATCCAAGCTATTGAC             304

 P  L  K  Y  E  V  M  T  I  G  I  A  P  T  M  D  W  Y  W  Y
CCGCTGAAATATGAAGTAATGACCATTGGCATCGCACCAACAATGGATTGGTATTGGTAT             364

 Q  G  N  L  A  N  V  R  N  D  T  W  L  E  D  H  K  N  C  H
CAAGGAAACCTCGCGAATGTTCGCAATGATACTTGGCTAGAAGATCACAAAAACTGTCAC             424

 Q  L  T  F  S  S  Q  G  F  I  L  G  E  K  R  I  V  P  D  V
CAGCTGACTTTTTCTAGCCAAGGATTTATATTAGGAGAAAAACGAATCGTCCCTGATGTC             484

 L  F  P  V  L  H  G  K  Y  G  E  D  G  C  I  Q  G  L  L  E
CTCTTTCCAGTCTTGCATGGGAAGTATGGCGAGGATGGCTGTATCCAAGGACTGCTTGAA             544

 L  M  N  L  P  Y  V  G  C  H  V  A  A  S  A  L  C  M  N  K
CTAATGAACCTGCCTTATGTTGGTTGCCATGTCGCTGCCTCCGCATTATGTATGAACAAA             604

 W  L  L  H  Q  L  A  D  T  M  G  I  A  S  A  P  T  L  L  L
TGGCTCTTGCATCAACTTGCTGATACCATGGGAATCGCTAGTGCTCCCACTTTGCTTTTA             664

 S  R  Y  E  N  D  P  A  T  I  D  R  F  I  Q  D  H  G  F  P
TCCCGCTATGAAAACGATCCTGCCACAATCGATCGTTTTATTCAAGACCATGGATTCCCG             724
```

```
 I  F  I  K  P  N  E  A  G  S  S  K  G  I  T  K  V  T  D  K
ATCTTTATCAAGCCGAATGAAGCCGGTTCTTCAAAAGGGATCACAAAAGTAACTGACAAA     784

 T  A  L  Q  S  A  L  T  T  A  F  A  Y  G  S  T  V  L  I  Q
ACAGCGCTCCAATCTGCATTAACGACTGCTTTTGCTTACGGTTCTACTGTGTTGATCCAA     844

 K  A  I  A  G  I  E  I  G  C  G  I  L  G  N  E  Q  L  T  I
AAGGCGATAGCGGGTATTGAAATTGGCTGCGGCATCTTAGGAAATGAGCAATTGACGATT     904

 G  A  C  D  A  I  S  L  V  D  G  F  F  D  F  E  E  K  Y  Q
GGTGCTTGTGATGCGATTTCTCTTGTCGACGGTTTTTTTGATTTTGAAGAGAAATACCAA     964

 L  I  S  A  T  I  T  V  P  A  P  L  P  L  A  L  E  S  Q  I
TTAATCAGCGCCACGATCACTGTCCCAGCACCATTGCCTCTCGCGCTTGAATCACAGATC     1024

 K  E  Q  A  Q  L  L  Y  R  N  L  G  L  T  G  L  A  R  I  D
AAGGAGCAGGCACAGCTGCTTTATCGAAACTTGGGATTGACGGGTCTGGCTCGAATCGAT     1084

 F  F  V  T  N  Q  G  A  I  Y  L  N  E  I  N  T  M  P  G  F
TTTTTCGTCACCAATCAAGGAGCGATTTATTTAAACGAAATCAACACCATGCCGGGATTT     1144

 T  G  H  S  R  Y  P  A  M  M  A  E  V  G  L  S  Y  E  I  L
ACTGGGCACTCCCGCTACCCAGCTATGATGGCGGAAGTCGGGTTATCCTACGAAATATTA     1204

 V  E  Q  L  E  A  L  A  E  E  D  K  R  *
GTAGAGCAATTGATTGCACTGGCAGAGGAGGACAAACGATGAACACATTACAATTGATCAATA  1267

AAAACCATCCATTGAAAAAAAATCAAGAGCCCCCGCACTTAGTGCTAGCTCCTTTTAGCGATCACGATG  1336

TTTACCTGCAG                                                      1347
    PstI
```